



Titre: Estimation de l'incertitude sur les flux d'inventaire du cycle de vie -
modélisation et développement de facteurs empiriques pour
l'approche pedigree
Title:

Auteur: Stéphanie Muller
Author:

Date: 2015

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Muller, S. (2015). Estimation de l'incertitude sur les flux d'inventaire du cycle de
vie - modélisation et développement de facteurs empiriques pour l'approche
Citation: pedigree [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie.
<https://publications.polymtl.ca/1965/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie:
PolyPublie URL: <https://publications.polymtl.ca/1965/>

**Directeurs de
recherche:** Réjean Samson, & Pascal Lesage
Advisors:

Programme: Génie chimique
Program:

UNIVERSITÉ DE MONTRÉAL

ESTIMATION DE L'INCERTITUDE SUR LES FLUX D'INVENTAIRE DU CYCLE DE VIE
– MODÉLISATION ET DÉVELOPPEMENT DE FACTEURS EMPIRIQUES POUR
L'APPROCHE PEDIGREE

STÉPHANIE MULLER

DÉPARTEMENT DE GÉNIE CHIMIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIAE DOCTOR
(GÉNIE CHIMIQUE)

DÉCEMBRE 2015

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

ESTIMATION DE L'INCERTITUDE SUR LES FLUX D'INVENTAIRE DU CYCLE DE VIE
– MODÉLISATION ET DÉVELOPPEMENT DE FACTEURS EMPIRIQUES POUR
L'APPROCHE PEDIGREE

présentée par : MULLER Stéphanie

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. LEGROS Robert, Ph. D., président

M. SAMSON Réjean, Ph. D., membre et directeur de recherche

M. LESAGE Pascal, Ph. D., membre et codirecteur de recherche

Mme GAUDREAUULT Caroline, Ph. D., membre

M. NORRIS Gregory A., Ph. D., membre

DÉDICACE

À tous ceux qui sont acteurs plutôt que spectateurs

«Des vertiges d'apprendre à comprendre » Jean-Jacques Goldman

« - What if you haven't the data? – Then we shall proceed directly proceed to the brandy and cigars. » Sir Arthur Conan Doyle

« We balance probabilities and choose the most likely. It is the scientific use of imagination. »

Sir Arthur Conan Doyle

« If our appreciation of uncertainty improves, our predictions can get better too. This is the “prediction paradox”: the more humility we have about our hability to make decisions, the more successful we can be in planning for the future. » Rachel Maddow

« A decision made without taking uncertainty into account is barely worth calling a decision. »

Richard Wilson

« La vérité ce n'est pas le certain, et l'incertain ce n'est pas l'ignorance. » Ilya Prigogine

REMERCIEMENTS

À tous ceux qui m'ont entourée et soutenue pendant cette thèse, merci!

Merci Réjean pour ton accueil au sein du CIRAIG et de la Chaire, pour ton support, tes conseils et ta disponibilité.

Merci Pascal pour ton encadrement tout au long de ce projet, pour ta disponibilité, pour tes réponses face à mes interrogations; pour tes questions, tes commentaires, ton regard critique qui m'ont aidée à avancer tout au long de ce doctorat. Merci également pour m'avoir offert ce projet que je continue encore à apprécier.

Thanks Chris for your warm welcome in Zürich, for your time, for all the exchanges and special thanks for making me less scared by python. Viele Dank Andreas für Ihre Begleitung in der erste Phase des "Pedigree Projekt". Also thank to Bo Weidema for bringing your vision to one of the objective of this project.

Merci à la grande famille du CIRAIG! Votre disponibilité, votre bonne humeur, les échanges qui ont lieu et l'ambiance qui se crée autour de vous, les individus qui composent cette famille en font un lieu de travail privilégié. Un merci spécial à Valérie et Isabelle pour m'avoir ouvert la porte du CIRODD ainsi qu'à Annie pour m'avoir fait confiance et permis d'apporter ma patte en GCH2220 puis GCH1220.

Merci aux partenaires de la Chaire internationale du cycle de vie ainsi qu'aux ministères engagés dans le projet de la base de données d'inventaire québécoise pour leur support financier.

Aux étudiants avec lesquels j'ai partagé mes bureaux successifs, merci! Merci d'avoir enduré mes quelques pertes de sang froid, mes réactions qui sortent parfois de nulle part et mes 90 minutes hebdomadaires à supporter le Racing. Une mention spéciale à Ariane, François, Viêt et Vincent pour l'inauguration des « bureaux du 3^{ème} », pour les rires et la chanson de Céline. Merci Laure, Leyla et Yohan d'avoir perpétué cette tradition d'ambiance bon enfant dans ce bureau du 3^{ème}. Yohan, merci pour tes conseils et tes explications sur python. Merci Gaël, Laure et Viêt pour vos temps de cerveau disponible, pour avoir apporté votre regard externe et vos recommandations et conseils lors de mes moments de doute. Et parce qu'au-delà d'un endroit de recherche, ces bureaux sont également des endroits d'échanges et de partage - certes sur la connaissance ACV mais pas que – et ces partages ont joué un rôle dans l'épanouissement de la

rédatrice de cette thèse. Cela va des partages culinaires (François, Laure, Pablo merci de m'avoir nourrie, je n'ai pas réussi à déterminer le meilleur cuisinier), aux partages musicaux et de la sphère internet (merci Constant), au partage de chocolat (thanks Hassana for taking care of me during the last weeks of writing) en passant par le partage d'apéro ou de poutines. Merci à vous tous pour tout ça!

En dehors de la sphère du CIRAIG et de l'ACV je me suis également sentie soutenue dans cette aventure par des personnes qui, elles aussi, méritent un merci ici.

À tous ceux avec lesquels je me suis évadée sur les planches, merci! Dana, Fannie, Guillaume, Mathieu, Menel et Victoria merci d'avoir joué un si grand rôle dans ma vie ces dernières années. Merci de m'avoir donné l'opportunité de grandir à vos côtés! Merci Émilie pour ces discussions très souvent autour de théières. François, Shirley et Viêt simplement merci d'être là que ce soit avec captain, avec du chocolat, avec des poulets ou des dindes, avec des voitures de location, avec des meeples (surtout les verts et les bleus), avec un plateau de jeu 7 Wonders, avec des discussions passionnées sur notre système actuel ou juste avec un jeu de mot. Merci pour tous ces moments d'échanges et de partage, d'accepter mes blagues et le fait que j'aime m'asseoir à terre!

Merci à vous de l'autre côté de l'Atlantique. Blandine, Boris, Damien, Jean-Brice, Line, Matthieu, Mélanie et Vincent merci pour ces moments de retrouvailles toujours organisés à « l'arrache » mais toujours plein de charmes. Merci à Arnaud pour le traditionnel vin chaud, les Edelweiss et les discussions qui les accompagnent (par contre je n'ai toujours pas de réponse à ta question).

Enfin un immense merci à ma famille : Maman, Papa, Nico et Élise, merci pour votre soutien et votre présence (même si elle a souvent été virtuelle ces derniers temps), pour accepter mes réponses à vos questions qui sont souvent « je ne sais pas » et pour, malgré ça, me faire confiance dans mes choix.

RÉSUMÉ

L'analyse du cycle de vie (ACV) est un outil d'aide à la décision qui permet d'évaluer les impacts environnementaux potentiels d'un produit, service ou système sur l'entièreté de son cycle de vie : de l'acquisition des matières premières à la gestion de la fin de vie. Cet outil permet de comparer les performances environnementales de différents systèmes ou celles des phases du cycle de vie d'un même produit. Cette connaissance permet aux politiques, au secteur privé ou aux consommateurs de prendre des décisions par rapport aux systèmes étudiés.

En tant qu'outil d'aide à la décision basé sur la modélisation, l'ACV considère les incertitudes dues aux modèles employés et à leurs entrants afin d'améliorer le degré de confiance à accorder aux résultats et aux prises de décisions qui en découlent. La typologie d'incertitude la plus documentée est celle touchant les paramètres en inventaire du cycle de vie (ICV – étape de l'ACV permettant de comptabiliser l'ensemble des échanges entre un système et l'environnement). Ces paramètres sont à la fois incertains – on ne connaît pas leur « vraie » valeur – et variables – leur valeur change selon le temps, l'espace ou le type d'individus.

Aucune méthode pour traiter l'incertitude et la variabilité sur les paramètres d'ICV n'est actuellement utilisée de manière universelle. Seule la base de données d'ICV ecoinvent propose une approche semi-quantitative de gestion de l'incertitude qui prend en compte l'ensemble des données d'arrière-plan, d'ores et déjà modélisées dans la base de données, pour le calcul d'incertitude. Cette approche, ici appelée « approche pedigree », combine une approche statistique et l'utilisation d'une matrice pedigree pour combiner deux typologies d'incertitude. Cette approche quantifie donc 1) l'incertitude de base qui représente l'erreur stochastique ou l'inexactitude et 2) l'incertitude additionnelle qui témoigne de l'utilisation d'une donnée imparfaite à la fois pour le contexte de l'étude et par rapport à sa source.

Cette approche pedigree, telle qu'appliquée dans la version 2 de la base de données ecoinvent, souffre de certaines limites : l'approche n'est valable que pour des données modélisées avec leur incertitude par une distribution log-normale, les différents facteurs d'incertitude de base et additionnelle ont été développés par jugements et ne possèdent donc pas d'assise empirique, de plus les facteurs d'incertitude sont les mêmes quel que soit le secteur industriel concerné.

Cette thèse propose d'aborder ces limites en répondant à l'objectif global de développer et légitimer par une assise scientifique la structure de gestion semi-quantitative de l'incertitude sur les données d'inventaire qu'est l'approche pedigree. Pour ce faire, les trois objectifs spécifiques suivants sont traités : 1) légitimer l'usage de la distribution log-normale comme distribution par défaut pour modéliser une donnée d'inventaire avec son incertitude; 2) déterminer comment l'approche pedigree peut être appliquée à des distributions autres que log-normales; 3) développer une méthode d'analyse de données permettant de déterminer les facteurs d'incertitudes, et ce, par secteurs d'activités.

En répondant à ces objectifs spécifiques, cette thèse apporte les contributions qui suivent à la modélisation de l'incertitude sur les données d'ICV. Tout d'abord, l'analyse de l'effet du choix d'une distribution par défaut pour modéliser une donnée et son incertitude montre que la distribution par défaut utilisée a une incidence sur les résultats d'une ACV quand on analyse les systèmes de manière individuelle. Néanmoins, dans le cas de comparaison de systèmes de produits, le choix de la distribution par défaut ne modifie pas les résultats d'une comparaison. Compte tenu de ces résultats, la distribution log-normale peut donc être utilisée par défaut pour modéliser une donnée et son incertitude si aucune information ne permettant de définir spécifiquement la distribution de la donnée n'est disponible. Le développement d'équations de compilation d'incertitude de base et d'incertitude additionnelle pour l'ensemble des distributions présentes dans la base de données ecoinvent pour modéliser une donnée et son incertitude permet aux praticiens et aux développeurs de données d'utiliser la distribution qui modélise le mieux la donnée. Finalement, par l'utilisation du théorème de Bayes, une méthodologie permettant de mettre à jour les facteurs d'incertitude est développée. Cette méthodologie s'applique si de nouvelles informations sur ces derniers sont disponibles. L'application de cette méthodologie dans le cadre de cette thèse permet l'obtention de facteurs d'incertitude empiriques et mis à jour, et ce, par secteurs industriels. Ces facteurs sont, pour la plupart des cas, plus élevés que ceux obtenus par jugements d'experts.

Les développements mis en œuvre pour atteindre ces objectifs sont basés sur un certain nombre d'hypothèses qu'il s'agit de garder à l'esprit lors de l'utilisation de ces résultats. Malgré ces hypothèses, cette thèse propose une justification complète de l'approche pedigree. De plus, ces travaux ont de l'intérêt que si ces derniers sont utilisés par les praticiens ou par les développeurs de données pour lesquels ces résultats s'appliquent. En ce sens, des discussions ont été engagées

pour intégrer les facteurs d'incertitude développés dans la base de données ecoinvent. L'approche développée pourrait facilement être intégrée dans d'autres bases de données.

Finalement, cette thèse participe à l'amélioration de la modélisation et donc de la connaissance de l'incertitude sur les données d'ICV. Connaissance nécessaire pour pouvoir considérer de manière systématique l'incertitude en ACV et la considérer comme une source de connaissance en soi sur les résultats d'une ACV plutôt qu'être considérée comme une faiblesse.

ABSTRACT

Life cycle assessment (LCA) is a decision support tool that allows evaluating the potential environmental impacts of a product, service or system from the perspective of its whole life cycle: from resource extraction to end-of-life. LCA aims to compare the environmental performances of different systems or of the different life cycle stages of a same product. This knowledge allows legislators, the private sector or consumers to consider environmental aspects in their decisions regarding the assessed systems.

As a decision support tool that relies on modelling, uncertainties are an intrinsic part of LCA models. The uncertainty of models and their inputs can be evaluated during a LCA. This evaluation can help improve the confidence in the results and in the decision the LCA results support. The most documented typology of uncertainty is uncertainty on life cycle inventory (LCI) parameters (LCI is the step of the LCA that aims to quantify all the exchanges between the system and the environment). These parameters are both uncertain – we don't know their “true” value – and variable – their value changes according to time, space and type of individual.

There is presently no universal method to address LCI parameter uncertainty and variability. Only the ecoinvent LCI database offers a semi-quantitative approach to manage and quantify uncertainty on all the background data modelled in the database. This approach, called “the pedigree approach”, is based on the use of statistics and a pedigree matrix to combine two typologies of uncertainty. This approach quantifies 1) the basic uncertainty that represents the stochastic error ;and 2) the additional uncertainty that is due to the use of an imperfect data, imperfect according to the context study and according to how the datum was obtained.

This pedigree approach, as applied in ecoinvent v2 has several limitations: the approach can only be applied if a datum with its uncertainty is modelled by a lognormal distribution, the uncertainty factors for both the basic and the additional uncertainties rely on experts' judgments rather than on empirical basis and the additional factors are unique whatever industrial sector.

This thesis aims to address these limits by developing and legitimating, on a scientific basis, the semi-quantitative uncertainty management tool on uncertainty parameters that is the pedigree approach. In order to meet this global objective, three specific objectives are addressed: 1) legitimate the use of the lognormal distribution as a default distribution to model a parameter

with its uncertainty; 2) determine how the pedigree approach can be applied to distributions other than lognormal; and 3) develop a data assessment structure that permits to determine empirically-based uncertainty factors and that, by activity sectors.

By meeting these specific goals, this thesis makes several contributions to the modelling of uncertainty of LCI data. First of all, the analysis of the effects of the choice of default distribution to represent uncertainty of background LCI data shows that, while the choice can significantly affect the LCA results when product systems are assessed individually, it has no material effect on products system comparisons. Giving these results, the lognormal distribution can be used as a default distribution to model a datum and its uncertainty and that, if no information to specifically define the datum's distribution is available. The development of equations that combine basic uncertainty and additional uncertainty for all distributions in the ecoinvent database to model a datum and its uncertainty allows the practitioner and the data provider to choose the distribution that best fits the assessed datum. Lastly, with the use of Bayesian inference, a methodology that aims to refine uncertainty factors is developed. Applying this methodology allows to refine uncertainty factors whenever new data are accessible. This methodology is applied in the context of this thesis, which led to the development of empirically based, refined and sector-specific uncertainty factors. These factors are mostly higher than the ones presently used in ecoinvent.

The developments related to the achievements of the goals are based on some assumptions that need to be considered when the results are used. Despite these assumptions, this thesis brings a complete justification of the use of the pedigree approach. This work is of interest only if the results are used by the practitioners or the data developers who need them. Discussions are engaged with the ecoinvent centre to integrate the developed uncertainty factors into the database. The approach could also easily be used with other databases.

Finally, this thesis contributes to the improvement of uncertainty modeling and to the knowledge on uncertainty on LCI parameters. This knowledge is necessary to systematically consider uncertainty in LCA and to consider uncertainty on results as a source of information on these results rather than a source of weakness.

TABLE DES MATIÈRES

DÉDICACE.....	III
RÉSUMÉ.....	VII
ABSTRACT	X
TABLE DES MATIÈRES	XII
LISTE DES TABLEAUX.....	XVI
LISTE DES FIGURES.....	XVIII
LISTE DES SIGLES ET ABRÉVIATIONS	XXI
LISTE DES ANNEXES.....	XXII
CHAPITRE 1 INTRODUCTION.....	1
1.1 Mise en contexte.....	1
1.2 La problématique de l'incertitude dans les prises de décisions basées sur l'ACV	5
CHAPITRE 2 REVUE DE LA LITTÉRATURE.....	7
2.1 Concepts généraux autour de l'incertitude.....	7
2.1.1 Incertitudes, variabilités et qualité	7
2.1.2 La prise en compte de l'incertitude en sciences appliquées.....	8
2.2 La prise en compte de l'incertitude en ACV	10
2.3 La gestion de l'incertitude en ICV	16
2.3.1 Typologies et sources d'incertitude en ICV – le cas particulier de l'incertitude sur les paramètres	16
2.3.2 Représentation de l'incertitude et de la qualité des données en ICV	18
2.3.3 Méthodes de propagation des incertitudes sur les paramètres	38
2.3.4 La gestion de l'incertitude dans les bases de données d'inventaire	39

2.4	Une approche de gestion semi-quantitative de l'incertitude appliquée à la base de données ecoinvent : l'approche pedigree	41
2.4.1	Le système NUSAP.....	41
2.4.2	Application de l'approche pedigree pour la gestion de l'incertitude dans la base de données ecoinvent	42
2.5	Problématiques issues de la revue de littérature.....	49
2.5.1	La nécessité d'une approche permettant la gestion de l'incertitude sur l'ensemble des données utilisées en ICV	49
2.5.2	Les limites actuelles de l'approche pedigree.....	50
CHAPITRE 3	OBJECTIFS DU PROJET ET MÉTHODOLOGIE GÉNÉRALE	52
3.1	La première phase du « projet pedigree »	52
3.2	Définition des objectifs et du projet de recherche	52
3.3	Méthodologie générale.....	53
3.3.1	Légitimer l'usage de la distribution log-normale comme distribution par défaut.....	54
3.3.2	Application de l'approche pedigree aux autres types de distribution présentes dans ecoinvent v3	59
3.3.3	Développement de facteurs empiriques pouvant être mis à jour	61
CHAPITRE 4	ARTICLE 1: EFFECTS OF DISTRIBUTION CHOICE ON THE MODELLING OF LCI UNCERTAINTY – AN ASSESSMENT ON THE ECOINVENT V2.2 DATABASE.....	65
4.1	Présentation de l'article.....	65
4.2	Manuscrit.....	65
4.2.1	Introduction	65
4.2.2	Materials and methods	71
4.2.3	Results and discussion.....	76
4.2.4	Limits and recommendations	86

CHAPITRE 5	ARTICLE 2: THE APPLICATION OF THE PEDIGREE APPROACH TO THE DISTRIBUTIONS FORESEEN IN ECOINVENT V3	89
5.1	Présentation de l'article	89
5.2	Manuscrit.....	90
5.2.1	Introduction	90
5.2.2	Materials and methods	96
5.2.3	Results and discussion.....	102
5.2.4	Conclusions	108
CHAPITRE 6	ARTICLE 3: GIVING AN EMPIRICAL BASIS FOR UNCERTAINTY FACTORS USED IN GLOBAL LIFE CYCLE INVENTORY DATABASES: AN ALGORITHM TO UPDATE FACTORS USING NEW INFORMATION	109
6.1	Présentation de l'article	109
6.2	Manuscrit.....	109
6.2.1	Introduction	109
6.2.2	Methods.....	113
6.2.3	Results	123
6.2.4	Discussion and conclusion	131
CHAPITRE 7	RÉSULTATS COMPLÉMENTAIRES	134
7.1	Questions soulevées par l'obtention de nouveaux facteurs d'incertitude	134
7.2	Comparaison des analyses d'incertitude menées avant et à la suite de l'obtention de nouveaux facteurs.....	134
CHAPITRE 8	DISCUSSION GÉNÉRALE ET RECOMMANDATIONS	139
8.1	Atteinte des objectifs de la recherche.....	139
8.2	Les limites des résultats obtenus	141
8.2.1	Des hypothèses sur la modélisation des facteurs d'incertitude	141

8.2.2	La représentativité des données utilisées pour développer de nouveaux facteurs d'incertitude	144
8.2.3	Des facteurs d'incertitude additionnelle surestimés?	145
8.3	Recommandations	149
8.3.1	Recommandations pour de futures recherches	149
8.3.2	Comment intégrer les travaux dans les bases de données?	153
8.3.3	Comment utiliser les travaux ici développés?	154
CHAPITRE 9	CONCLUSION	156
9.1	Rappel des résultats et de l'originalité du projet	156
9.2	Intégration des résultats obtenus et futurs développements	157
9.3	Démocratiser la prise en compte de l'incertitude et de la variabilité en ACV	158
BIBLIOGRAPHIE	159
ANNEXES	167

LISTE DES TABLEAUX

Tableau 2-1 Exemples d'incertitude et de mode de gestion suivant leur typologie et les phases de l'ACV	13
Tableau 2-2 Exemples de critères utilisés en ICV pour coder la qualité d'une donnée.....	21
Tableau 2-3 PDF et représentation graphiques de différentes lois utilisées en ICV pour modéliser l'incertitude sur les données	26
Tableau 2-4 Principes de base d'arithmétique des intervalles	31
Tableau 2-5 Différences entre approches statistiques et bayésiennes.....	35
Tableau 2-6 Avantages et inconvénients des différentes méthodes de traitement de l'incertitude en ICV	37
Tableau 2-7 Matrice pedigree, traduite, appliquée dans la base de données ecoinvent et issue des travaux de Weidema et Wesnaes (1996)	43
Tableau 2-8 Tableau des facteurs d'incertitude de base appliqués par défaut dans la base de données ecoinvent (Weidema, et al. 2010)	46
Tableau 2-9 Tableau de conversion des scores pedigree en facteurs d'incertitude additionnelle.	48
Tableau 4-1 Characteristics of probability distribution and their use in LCI.....	68
Tableau 4-2 Determination of the definition parameters for the distribution used for the switch.	73
Tableau 4-3 Methodological scheme to assess the effects of the used distribution of product systems comparison.....	76
Tableau 4-4 Obtained trends in the comparison on the resulting distributions between the lognormal distribution used by default and each of the three switches. Ratios are calculated taking the lognormal distribution as a numerator.....	83
Tableau 5-1 Definition of the distributions foreseen in ecoinvent v3 and how to express the CV according to the definition parameters	93
Tableau 5-0-2 Correspondences between the normal and the lognormal distributions	99
Tableau 5-0-3 : Assumptions and resulting formulas for symmetric distributions.....	102

Tableau 5-4 : Assumptions and resulting formulas for asymmetric distributions	103
Tableau 5-5 : Definition parameters used to define the tested distribution that modelled a datum with its basic uncertainty and resulting calculated parameters when additional uncertainty is added. The basic and additional uncertainty combination is made using the proposed formulas foreseen in Tableau 5-0-3 and Tableau 5-4	105
Tableau 6-1 The three steps methodology used to develop updated uncertainty information.....	114
Tableau 6-2 Proposed pedigree matrix cell description for the completeness indicator.....	120
Tableau 6-3 Updated additional uncertainty factors obtained by using σ_{UF1} for the prior distribution (see Figure 6-2), the results for σ_{UF2} can be found in the Supporting information. The prior is theecoinvent value; L stands for the likelihood value and P for the posterior value.	127
Tableau 6-4 Updated additional uncertainty factors for the further technological correlation indicator for the manufacturing sector using a new data source	129
Tableau 6-5 Obtained posterior values for the basic uncertainty factors. The posterior 1 values were obtained using σ_{UF1} (see Figure 6-2) and the posterior 2 values using σ_{UF2}	130
Tableau 8-1 Résultats de l'analyse bayésienne par MCMC pour une fonction à priori uniforme et comparaison avec les résultats obtenus à l'article 3. La <i>fonction à posteriori1</i> représente la fonction obtenue à l'article 3 et la <i>fonction à posteriori2</i> représente celle obtenue dans le cas d'une loi à priori uniforme.	143

LISTE DES FIGURES

Figure 1-1	Modélisation d'un processus élémentaire avec l'ensemble de ses entrants et extrants.	3
Figure 1-2	Cadre de l'ACV selon la norme ISO 14040 :2006.....	4
Figure 2-1	Nature, niveaux et porteurs de l'incertitude au sein d'un système. Ces trois composantes peuvent servir à définir la typologie de l'incertitude. Adaptée de (Refsgaard et al., 2007; Swart et al., 2009; Wardekker et al., 2008; Warmink, et al., 2011).....	9
Figure 2-2	Évolution du nombre de publications par années – entre 1993 et 2014 – contenant dans leurs mots-clés les termes « incertitudes » et « analyse du cycle de vie ».....	12
Figure 2-3	Sources de données et caractéristiques influant sur leur qualité et incertitude. Adapté de (Leroy, 2009; Lloyd & Ries, 2007; Maurice et al., 2000; Weidema & Wesnæs, 1996)... ..	18
Figure 3-1	: Étapes de l'approche pedigree, limites associées et traitement de ces dernières pour répondre à l'objectif de recherche	55
Figure 3-2	: Schéma méthodologique général.....	56
Figure 4-1	Effect of the choice of the default distribution on the central tendency, expressed as the ratio between the median of the resulting distribution and the deterministic value and for each assessed default distribution used to model uncertainty of inputs	78
Figure 4-2	Effect of a switch in choice of default distribution on the central tendency, expressed as the ratio between the median of the resulting distribution when an alternative default distribution is used and when the lognormal distribution is used as default.....	79
Figure 4-3	Effect of a switch in choice of default distribution on dispersion, expressed as the ratio between the standard deviation of the resulting distribution when an alternative distribution is used by default and when the lognormal distribution is used as default (logarithmic scale)	81
Figure 4-4	Skewness distribution γ_1 for each product system – on logarithmic scale	82
Figure 4-5	Resulting distribution for the climate change impact for the three switch and for the same product systems.....	83

Figure 4-6 Monte Carlo results for the difference between two product systems (P'_i and $\alpha P'_j$). Percentages indicate the probability for the result to be negative or positive.....	85
Figure 4-7 Ratio on the product system comparison, lognormal distribution as a denominator ...	86
Figure 5-1 : Illustration of the different distributions foreseen in Tableau 5-5.The histograms represent the results obtained through the Monte Carlo simulation.....	107
Figure 6-1 Illustration of the « Russian dolls » principle used to classify data into subgroups for the additional uncertainty	117
Figure 6-2 Assumptions used to apply the Bayes theorem – <i>Notations log. (logarithm), min. (minimum), max. (maximum)</i>	125
Figure 6-3 Additional uncertainty factors representation for the 5 pedigree indicators and the 4 pedigree scores. The value of Posterior 1 is obtained using σ_1 for the prior distribution, the value of Posterior 2 is obtained using σ_2 (see Figure 6-2)	126
Figure 6-4 Representation of the further technological correlation indicator for the manufacturing sector updated two times: once using the the master datatable used in this study, the second time by accessing a new data source from the INIES database.....	131
Figure 7-1 Représentation de la distribution du ratio des moyennes des distributions résultantes. Le ratio est effectué entre la moyenne de la distribution résultante obtenue pour le changement des facteurs d'incertitude et la moyenne de la distribution résultante initiale.	136
Figure 7-2 Représentation de la distribution du ratio des écarts-type des distributions résultantes. Le ratio est effectué entre l'écart-type de la distribution résultante obtenue pour le changement des facteurs d'incertitude et l'écart-type de la distribution résultante initiale.	137
Figure 7-3 Représentation des distributions résultantes de l'impact sur les changements climatiques pour deux systèmes de produits de la base de données ecoinvent v2.2.....	138
Figure 8-1 Représentation des distributions à priori (<i>prior</i>) uniforme et à posteriori (<i>posterior</i>) ainsi que de la fonction de vraisemblance (<i>likelihood</i>) pour le score 5 de l'indicateur justesse	144
Figure 8-2 : Pourcentages des sous-secteurs NAICS définis au niveau 6 représentés dans l'analyse de données pour chaque secteur industriel spécifique	145

Figure 8-3 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de justesse (R) et de corrélation technologique (FTC) sont étudiés simultanément.....	147
Figure 8-4 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de corrélation géographique (GC) et de corrélation technologique (FTC) sont étudiés simultanément	148
Figure 8-5 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de corrélation temporelle (TC) et de corrélation technologique (FTC) sont étudiés simultanément	149
Figure 9-1 : Schéma méthodologique général de mise en œuvre d'une ACV – adapté de ISO 14040 (ISO, 2006a)	167

LISTE DES SIGLES ET ABRÉVIATIONS

ACV	Analyse du cycle de vie
CV	Coefficient de variation – rapport entre l'écart-type arithmétique et la moyenne arithmétique
DQI	<i>Data quality indicator</i> – Indicateur de qualité des données
ÉICV	Évaluation des impacts du cycle de vie
ELCD	<i>European reference life cycle database</i> – Base de données européenne de référence du cycle de vie
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
GSD	<i>Geometric standard deviation</i> - écart-type géométrique
ICV	Inventaire du cycle de vie
ISIC	<i>International Standard Industrial Classification</i> – Classification internationale des secteurs industriels
MCMC	<i>Markov Chains Monte Carlo</i> - Simulations de Monte-Carlo par chaînes de Markov
NAICS	<i>North American Industrial Classification System</i> – Classification Nord-américaine des secteurs industriels
NUSAP	<i>Numerical Unit Spread Assessment Pedigree</i> – Système de gestion et de communication de l'incertitude et de la qualité
ODD	Objectifs de développement durable
OMD	Objectifs du millénaire pour le développement
PDF	<i>Probability density function</i> - Fonction de densité de probabilités
PNUE	Programme des Nations-Unies pour l'environnement
SETAC	Société de chimie et de toxicité environnementale

LISTE DES ANNEXES

Annexe A – Des données autonomes aux impacts potentiels, le modèle ACV	167
Annexe B – Matériel supplémentaire publié avec l'article 1 (Chapitre 4).....	171
Annexe C – Matériel supplémentaire publié avec l'article 2 (Chapitre 5).....	180
Annexe D – Matériel supplémentaire publié avec l'article 3 (Chapitre 6)	184

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

Depuis notamment la création du Club de Rome en 1968 et la parution de *Halte à la croissance* en 1972, des prises de conscience de la part d'organisations et de la société civile sont nées autour de l'idée que le développement vers un bien-être social et économique ne peut se faire en négligeant les limitations en termes de ressources naturelles et des capacités finies de l'écosystème. Afin de limiter les impacts des activités humaines sur les systèmes naturels, mais également sur la pauvreté et les inégalités sociales, certains en appellent à un changement de paradigmes dans nos façons d'aborder le développement et de mettre en œuvre les principes du développement durable (UNEP, 2015).

Face à cette prise de conscience, bon nombre d'organisations, d'instances gouvernementales et intergouvernementales, d'entreprises mais également des citoyens se sont donc appropriés le concept de développement durable et en ont multiplié les définitions et les interprétations pour l'appliquer dans leur champ d'action parfois avec des perceptions différentes, plus ou moins complètes, du sujet (Lélé, 1991). En 1986, le rapport Brundtland présentait le développement durable de la sorte (Brundtland & CMED, 1987) :

« Le développement durable n'est pas un état d'équilibre mais plutôt un processus de changement dans lequel l'exploitation des ressources, le choix des investissements, l'orientation du développement technologique ainsi que le changement institutionnel sont déterminés en fonction tant des besoins qu'actuels qu'à venir. »

Qui dit processus de changement dans notre manière d'aborder notre développement, qu'il soit social ou économique, dit mise en place de plans d'actions pour initier et concrétiser les changements, et dit donc processus de prises de décisions afin de mettre en œuvre le « meilleur » des plans d'actions.

Dans le domaine environnemental, afin de faire des choix, de planifier des actions et d'émettre des politiques publiques pour limiter l'exploitation des ressources ou encore le stress sur les écosystèmes et la santé humaine, bon nombre d'outils servant à l'aide à la prise de décisions existent. Parmi ces derniers on pourra citer en exemple les outils d'aide multicritères à la décision appliqués au contexte environnemental, les outils de l'écoconception, les analyses de risque ou

encore les outils issus de la pensée cycle de vie telles que les empreintes eau ou carbone ou encore les analyses du cycle de vie (ACV).

L'ACV est un outil de gestion environnemental qui permet d'évaluer les impacts potentiels d'un produit, service ou système en se basant sur la fonction de l'entité à l'étude, et ce, sur l'entièreté de son cycle de vie : de l'acquisition des matières premières à sa gestion en fin de vie. Cet outil permet de comparer les performances environnementales de différents systèmes (possédant la même base fonctionnelle) ou encore celles des phases du cycle de vie d'un même produit. Cette connaissance des performances environnementales relatives permet de prendre des décisions par rapport au produit ou au service que ce soit pour son développement ou son amélioration, pour mettre en place des politiques publiques ou encore à des fins de marketing (ISO, 2006a). Actuellement, l'ACV est régi par deux normes internationales, la norme ISO 14040 :2006 qui définit les principes et les cadres d'une ACV et la norme ISO 14044 :2006 qui, quant à elle, fixe les exigences et les lignes directrices. Ainsi, selon ces normes ISO qui fixent le cadre méthodologique de mise en œuvre d'une ACV, une étude ACV est un processus itératif mené selon quatre grandes phases décrites ci-après et telles que présentées à la Figure 1-2.

- **La définition des objectifs et du champ de l'étude** permet de définir le problème et le contexte du ou des systèmes à l'étude en répondant aux questions « Pourquoi? Pour qui? Quoi? Quand? Comment? Où? ». Le système de produits est le système qui représente le cycle de vie d'un produit. Tout comme l'ensemble des systèmes, il est composé d'un certain nombre d'éléments ayant des interactions dynamiques et permettant de répondre à un objectif commun : ici la fonction principale. L'unité fonctionnelle est aussi définie au sein de cette phase, celle-ci permet de quantifier la fonction du système, c'est sur cette base que les résultats seront comparés et que l'ensemble des entrants et des extrants du système avec son environnement vont être rapportés. Le système de produits est composé de processus élémentaires qui témoignent des activités ayant lieu au sein du système pour créer la fonction.
- **L'inventaire du cycle de vie (ICV)** vise à quantifier l'ensemble des entrants et des extrants – les flux - de chaque processus élémentaire du système relativement à l'unité fonctionnelle précédemment définie. Ces flux sont de deux sortes, ils peuvent relier le système à l'environnement et on parlera alors de flux élémentaire (p.ex. émission de CO₂); ils peuvent relier les différents processus du système entre eux et on parlera de flux

économiques (p.ex. quantité d'électricité). La Figure 6-1 représente un processus élémentaire avec les différentes typologies de flux entrants et sortants qui peuvent s'y rattacher. Une quantité non négligeable de données doit donc être collectée et compilée afin de mener cette étape d'ICV. Néanmoins, un certain nombre de processus élémentaires et leurs flux associés sont d'ores et déjà modélisés et sont disponibles dans les bases de données d'inventaire. Ces processus modélisés ne sont généralement pas spécifiques à une entreprise ou un produit en particulier, mais sont des processus génériques à un secteur d'activités ou à une technologie et à une aire géographique (exemple de type de processus disponible dans une base de données d'inventaire : *Electricity, high voltage {CA-ON} electricity production, hydro run-of-river* qui témoigne de la production d'électricité à partir d'une centrale au fil de l'eau en Ontario).

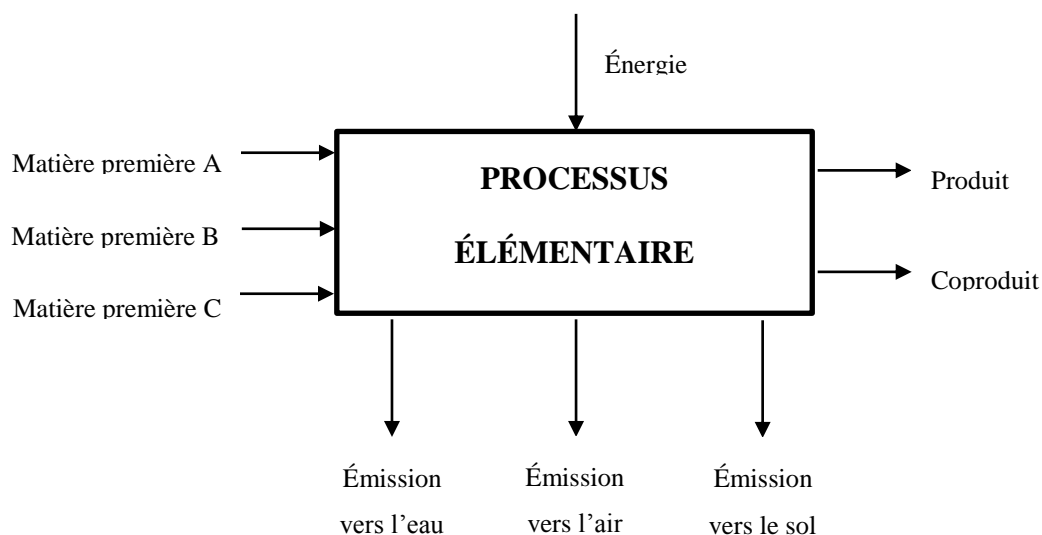


Figure 1-1 Modélisation d'un processus élémentaire avec l'ensemble de ses entrants et extrants

- **La phase d'évaluation des impacts du cycle de vie (ÉICV)** permet de relier chacun des flux élémentaires précédemment quantifiés à leur impact potentiel sur l'environnement. Plusieurs méthodologies d'ÉICV existent actuellement, néanmoins trois aires de protections communes peuvent être identifiées dans chacune de ces méthodes : la santé humaine, la qualité des écosystèmes et les ressources (J. R. C. 2010a).

- **La phase d'interprétation** permet, quant à elle, d'interpréter et de conclure par rapport aux résultats obtenus mais également d'évaluer les incertitudes qui émanent de chacune des phases précédentes (Jolliet, 2005).

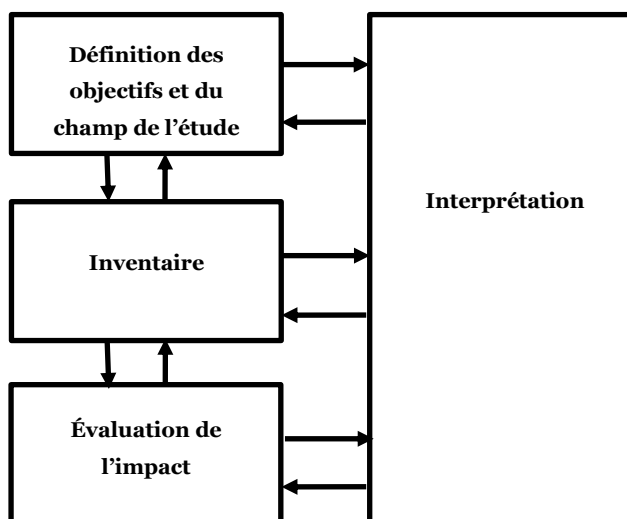


Figure 1-2 Cadre de l'ACV selon la norme ISO 14040 :2006

L'annexe A présente de manière plus détaillée l'ensemble des phases décrites ci-dessus en mettant l'accent sur la phase d'ICV.

Dans les dernières années, le volet social de l'ACV environnementale a été développé. Le premier guide méthodologique émis par la *Life cycle initiative* du Programme des Nations-Unies pour l'environnement (PNUE) et de la Société de chimie et toxicité environnementale (SETAC) date de 2009 (UNEP, 2009). De son côté, l'analyse du coût du cycle de vie est une technique développée bien avant les aspects environnemental et social. L'analyse des coûts permet de choisir l'option la plus efficiente en termes de coût (combinant coûts direct et indirect) en considérant l'ensemble de son cycle de vie. L'ACV environnementale peut donc être un outil d'aide à la décision en soi ou s'intégrer dans des analyses de la durabilité du cycle de vie où les volets cycle de vie des coûts et ACV sociale sont également pris en compte.

1.2 La problématique de l'incertitude dans les prises de décisions basées sur l'ACV

L'incertitude est souvent caractérisée par un savoir insuffisant, un manque de connaissances autour d'une situation donnée. Plus précisément, quantifier l'incertitude pourrait revenir à quantifier le degré de confiance que le décideur attribue à un résultat probable (Refsgaard et al., 2007). Cette absence de connaissances ou la mauvaise prise en compte de cette dernière peut entraîner la mise en application de décisions politiques imparfaites ou l'élaboration de stratégies erronées qui peuvent avoir *in fine* un impact négatif (Wardekker et al., 2008).

Aujourd'hui encore, les scientifiques, quel que soit leur domaine d'activité, sont hésitants à considérer et communiquer autour de l'incertitude. En effet, certains considèrent que les parties prenantes ne comprennent pas entièrement les enjeux liés à l'incertitude et que sa communication entraînerait une perte de crédibilité autour des résultats obtenus (Hansen, 2012; Wardekker et al., 2008). Néanmoins, comme mentionné précédemment, la prise en compte des incertitudes permet au décideur d'élaborer sa stratégie avec un certain niveau de confiance. Elles sont donc importantes à considérer afin d'augmenter le niveau de confiance que l'on peut accorder au résultat obtenu et afin de prendre des décisions en reconnaissant les limites de ce dernier. La prise en compte des incertitudes doit donc se faire pour un outil d'aide à la décision tel que l'ACV.

En effet, l'ACV est une analyse basée sur la modélisation : modélisation du système de produits et de l'agencement des processus élémentaires; modélisation des processus élémentaires; modélisation des chaînes de cause à effets permettant de convertir les flux élémentaires en impacts potentiels. Des incertitudes inhérentes aux modèles employés, aux hypothèses qui les régissent et aux données permettant de faire tourner ces modèles sont donc présentes dans la mise en œuvre d'une ACV. Les résultats issus de la modélisation et servant à la prise de décisions sont donc incertains pour ces multiples raisons.

Ce projet de recherche vise à développer une approche systématique de prise en compte des incertitudes sur les données issues de bases de données d'inventaire du cycle de vie (ICV). Cette amélioration va permettre d'augmenter la qualité et la fiabilité des résultats d'incertitude liés à l'inventaire et ainsi de permettre aux praticiens et aux décideurs de faire des choix en connaissance de cause. Plus particulièrement, ce projet vise à développer la modélisation sur les

données et à développer des facteurs d'incertitudes mis à jour pour les données employées dans les bases de données d'inventaire. Après avoir discuté de la prise en compte des incertitudes en ACV et en ICV et des écueils de cette mise à jour, la méthodologie et les résultats en découlant seront présentés sous forme d'articles scientifiques. Une discussion autour des limites de la recherche et des recommandations, puis une conclusion clôtureront ce manuscrit.

CHAPITRE 2 REVUE DE LA LITTÉRATURE

2.1 Concepts généraux autour de l'incertitude

2.1.1 Incertitudes, variabilités et qualité

Une définition générale de l'**incertitude** serait la suivante : « Caractère de ce qui ne peut être déterminé, connu à l'avance. – Ce qui ne peut être établi avec exactitude, qui laisse place au doute » (Larousse, 1995a). Si la définition se veut ici universelle, suivant les domaines et les courants de pensées considérés, celle-ci peut varier et ainsi devenir un sujet plus subjectif (Refsgaard et al., 2007; Williams, et al., 2009). Si Keynes (1937) donne une définition rapide de l'incertitude « Simplement on ne sait pas », beaucoup d'auteurs sont moins catégoriques sur le sujet. Par exemple, pour Björklund (2002), il s'agit « du manque de connaissance sur la valeur réelle d'une donnée » pour le cas particulier de l'incertitude sur les valeurs quantitatives. L'incertitude peut également provenir du manque de confiance, du doute que l'on a sur un état de la connaissance tout comme sur une quantité donnée. Tandis que Hong et collègues (2010) considèrent que l'incertitude est la conséquence aléatoire d'un système, Ciroth et collègues (2004) lient l'incertitude à la réalité en soulignant le fait qu'une valeur mesurée diffère de la « vraie valeur » d'une donnée (*i.e.* la valeur théorique de cette donnée), et ce, de manière probabiliste. Si les définitions et les rapports à l'incertitude peuvent varier, de plus en plus de techniques permettant de la prendre en compte ont toutefois été développées, et ce, dans différents domaines : la métrologie, la climatologie, l'analyse du risque financier ou encore l'évaluation du risque nucléaire ne sont que des exemples parmi tant d'autres. La prise en compte et la détermination de l'incertitude dans chacun de ces domaines ont néanmoins comme objectif principal de donner des résultats pour lesquels un degré de confiance suffisant peut être apporté et sur lesquels la prise de décision peut être basée.

Bon nombre d'auteurs, et notamment dans le domaine de l'ACV, associent généralement l'incertitude à la **variabilité** qui est « le caractère de ce qui se présente de diverses manières, qui change de nature » (Larousse, 1995b). La variabilité est donc une caractéristique propre à un objet en lui-même, elle témoigne des variations inhérentes d'un paramètre selon l'espace, le temps ou encore les individus (Huijbregts, 1998) et est de par ce fait irréductible. On peut par contre exprimer la variabilité de manière plus précise en régionalisant, par exemple, les modèles

ou les données ou encore en changeant l'échelle de résolution d'une étude. L'incertitude, quant à elle, provient à la fois de l'objet considéré, de son contexte et de la manière dont l'objet est interprété dans ce contexte : elle peut être réduite par de nouvelles recherches sur l'objet en question.

Si la variabilité est inhérente au système à l'étude et si l'incertitude dépend à la fois du système à l'étude et de sa modélisation, le concept de **qualité** est, quant à lui, à relier directement au contexte dans lequel le système s'insère. Ainsi « la qualité est l'ensemble des caractéristiques d'une entité qui confère à celle-ci l'aptitude à satisfaire des besoins exprimés ou implicites » (ISO, 2005), que les besoins soient ceux du producteur ou du consommateur. Dans le cadre de l'ACV, la qualité d'un modèle ou d'un paramètre est donc à définir en fonction des exigences spécifiées dans la phase de définition des objectifs et du champ de l'étude (cf. Figure 1-2). Ainsi, dans sa définition, la qualité peut influencer la variabilité et l'incertitude, mais si, à l'inverse, un système possède une grande incertitude, cela ne veut pas forcément dire que la qualité de cette dernière est faible (Cooper & Kahn, 2012; Funtowicz & Ravetz, 1990).

2.1.2 La prise en compte de l'incertitude en sciences appliquées

La prise en compte systématique de l'incertitude en sciences appliquées est encore peu répandue, et ce, pour différentes raisons. Ceci est notamment dû à sa complexité de représentation et à la diversité dans ses formes. Néanmoins, comme le soulignent Funtowicz et Ravetz dans leur ouvrage *Uncertainty and quality in science for policy* (1990), cette prise en compte de l'incertitude est nécessaire en sciences puisque ce sont sur les résultats d'études scientifiques que sont basées les décisions politiques importantes.

Selon Swart et collègues (2009), qui ont mené une étude autour de la gestion de l'incertitude au sein du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat), et Weber (2012) la faible prise en compte de l'incertitude vient principalement de la diversité des typologies de l'incertitude, de la diversité des manières d'approcher chacune de ces typologies et de la diversité des ordres de grandeurs de l'incertitude au sein d'une même méthode. Ainsi la Figure 2-1 présente un recensement des formes génériques de l'incertitude qui peut être définie selon le trio : *Nature – Niveaux – Porteurs*.

Cette diversité de typologie, et ce, à différents niveaux, entraîne également une diversité dans les moyens de définir et d'évaluer l'impact de l'incertitude sur les différents sortants d'un système. Ces approches peuvent être génériques ou être propres à un domaine particulier ou peuvent encore être objectives ou subjectives. Refsgaard et collègues (2007) en proposent quatorze qu'il est possible de classer selon des méthodes purement qualitatives, semi-quantitatives ou quantitatives; chacune des approches étant appropriée pour des typologies d'incertitudes considérées.

Nature de l'incertitude	Niveaux de l'incertitude	Porteurs de l'incertitude au sein du système
<ul style="list-style-type: none"> • Incertitude <i>épistémique</i> ou <i>Inexactitude</i> Incertaineté provenant d'une mauvaise compréhension du système et pouvant être diminuée • Incertitude <i>stochastique</i> ou <i>Imprécision</i> Variabilité intrinsèque ne pouvant être diminuée 	<ul style="list-style-type: none"> • Incertitude <i>statistique</i> • Incertitude sur les <i>scénarios</i> • Incertitude <i>qualitative</i> • <i>Ignorance reconnue</i> 	<ul style="list-style-type: none"> • Incertitude sur le <i>contexte</i> et les <i>frontières</i> • Incertitude sur les <i>entrants</i> • Incertitude sur la <i>structure du modèle</i> • Incertitude sur les <i>paramètres</i> • Incertitude sur les <i>techniques de modélisation</i> • Incertitude sur la <i>sélection de données résultantes</i>

Figure 2-1 Nature, niveaux et porteurs de l'incertitude au sein d'un système. Ces trois composantes peuvent servir à définir la typologie de l'incertitude. Adaptée de (Refsgaard et al., 2007; Swart et al., 2009; Wardekker et al., 2008; Warmink, et al., 2011).

De par ces diversités dans les formes et les approches, les scientifiques sont hésitants à communiquer autour de l'incertitude. En effet, sa considération comme étant l'ensemble des formes d'ignorance autour d'un système fait en sorte qu'elle est perçue négativement, notamment du grand public (Swart et al., 2009). De plus, suivant le domaine dans lequel l'incertitude doit être gérée, la diversité des parties prenantes impliquées peut entraîner une mésentente sur le sujet ainsi que des risques de mauvaises interprétations. En effet, si l'incertitude est considérée comme le degré de confiance accordé à un résultat, suivant l'interprétation réalisée par la partie prenante, celle-ci peut considérer un résultat certain alors qu'il n'en est rien et vice-versa. Tel est le cas pour le GIEC, où la difficulté de communication autour de la gestion de l'incertitude provient de la complexité des sciences climatiques et des effets rebonds de la modification du climat, mais aussi du fait que le groupe est composé de diverses parties prenantes dont des scientifiques, des économistes, des spécialistes des sciences sociales et qu'ils doivent rendre des comptes à des politiques et au grand public. Pour faire face à ces difficultés et harmoniser le dialogue, des guides communs pour une gestion et une communication harmonisées de l'incertitude ont vu le jour au GIEC et sont bonifiés pour chacun des rapports. Les guides pour le cinquième rapport publié en 2015 soulignent notamment le fait de donner une expression qualitative de la confiance donnée en un résultat et une mesure quantifiée, de manière probabiliste, de l'incertitude sur ces résultats (Budescu, et al., 2012; Mastrandrea et al., 2011).

Il ne s'agit pas ici de faire une revue complète des moyens de gérer et de communiquer les incertitudes suivant les domaines. Pour chacun d'eux, les auteurs proposent plusieurs types d'approches se rapportant aux différentes typologies d'incertitudes et l'ACV n'est pas en reste dans cette diversité.

2.2 La prise en compte de l'incertitude en ACV

Tout comme les autres outils ou supports d'aide à la décision, les résultats d'une ACV perdent en crédibilité quand les incertitudes sont trop importantes ou non considérées, ou encore quand certains aspects ne sont pas pris en compte dans l'étude - comme le cas de la toxicité des métaux par exemple (Pizzol, et al., 2011a). Cette considération de l'incertitude devient néanmoins

nécessaire pour les praticiens et les décideurs qui basent leurs analyses, leurs décisions et leurs choix règlementaires sur les résultats d'une ACV. C'est le cas notamment pour le gouvernement du Québec qui, dans sa loi sur la qualité de l'environnement - section VII « La gestion des matières résiduelles », cite la conduite d'une ACV pour démontrer qu'un projet peut déroger au paragraphe 53.4.1 de la loi (Gouvernement du Québec, 2011).

Si la pensée cycle de vie est plus ancienne, l'ACV telle qu'elle est pratiquée actuellement date du début des années 90. Pourtant, ce n'est qu'au milieu des années 2000 que l'incertitude en ACV devient un sujet d'études et de discussions sur lequel les auteurs publient, comme le témoigne la Figure 2-2. C'est de par sa nature même d'outil basé sur la modélisation que les différentes phases d'une ACV possèdent de l'incertitude qui se propage jusqu'au résultat final. Les auteurs publiant dans le domaine de l'ACV et de l'incertitude les classent généralement en trois grandes familles que l'on pourrait classer dans les porteurs de l'incertitude (cf. Figure 2-1). On distingue ainsi l'incertitude sur les modèles, l'incertitude sur les scénarios et l'incertitude sur les paramètres (qui représentent ici l'ensemble des données quantitatives nécessaires à la réalisation d'une ACV) (Leroy, 2009; Lloyd & Ries, 2007; Reap et al., 2008). Le Tableau 2-1 donne des exemples sur la typologie de l'incertitude dépendamment de la phase de l'ACV dans laquelle on se situe et de ses modes de gestion dans la phase d'interprétation.

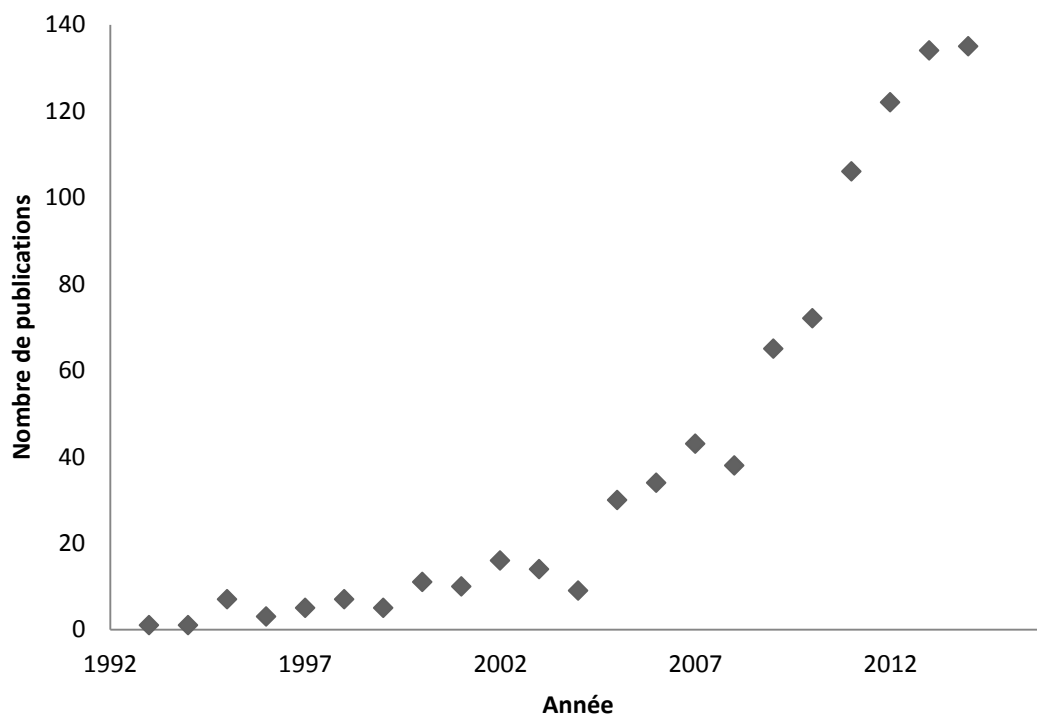


Figure 2-2 Évolution du nombre de publications par années – entre 1993 et 2014 – contenant dans leurs mots-clés les termes « incertitudes » et « analyse du cycle de vie » et obtenus par une recherche via « Engineering Village »¹.

¹ <http://www.engineeringvillage.com/>

Tableau 2-1 Exemples d'incertitude et de mode de gestion suivant leur typologie et les phases de l'ACV

	Incertitude sur les modèles	Incertitude sur les scénarios	Incertitude sur les paramètres
Définition des objectifs et du champ de l'étude	Détermination des frontières du système	Choix des scénarios de gestion des déchets	Paramètres-clés reliant l'unité fonctionnelle aux quantités de produits nécessaires
Inventaire	Choix des entrants et des sortants des processus élémentaires	Hypothèses posées pour réaliser l'inventaire (p.ex. transport)	Quantification des flux élémentaires et économiques
Évaluation des impacts	Choix de la chaîne de causes-à-effets	Hypothèses (p.ex. résolution spatiale pour la détermination des facteurs de caractérisation)	Valeurs quantitatives (EC ₅₀ ² sur lesquelles sont basés les calculs de facteurs de caractérisation p.ex.)
Interprétation – Gestion de	Raffiner le modèle (passe par	Analyse de scénarios : tester les	Approches quantitatives,

² EC₅₀: Concentration d'effet à 50%, concentration pour laquelle 50% de l'effet maximal est observé (p.ex. concentration pour laquelle 50% de la population testée présente une réponse au test)

	Incertitude sur les modèles	Incertitude sur les scénarios	Incertitude sur les paramètres
l'incertitude en fonction de la typologie	l'augmentation du nombre de paramètres et donc de l'incertitude sur ces derniers)	différentes hypothèses et leurs effets sur leur résultat final	qualitatives, semi- quantitatives. Propagation par approche analytique ou de Monte Carlo.

Comme en témoigne le Tableau 2-1, quelle que soit la phase de l'ACV, les différentes typologies de l'incertitude sont présentes. À cette diversité de typologies, il faut ajouter une diversité des ordres de grandeurs de l'incertitude quantifiée. Si l'on considère par exemple la propagation des incertitudes sur les processus élémentaires d'une base de données d'inventaire (ecoinvent v2.2.), les ordres de grandeur des coefficients de variation résultants (CV : rapport entre l'écart-type arithmétique et la moyenne arithmétique) varient de 10^{-4} à 10. Lorsqu'on se place dans la phase d'ÉICV, suivant les méthodologies employées on peut obtenir, dans le cas p.ex. de l'impact des métaux, des résultats variant de différents ordres de grandeurs (Pizzol et al., 2011a; Pizzol, et al., 2011b). Les facteurs de caractérisation pour certaines catégories d'impact peuvent eux-aussi varier de différents ordres de grandeurs (1 à 2 pour l'acidification p.ex.) (Roy, 2012). De par cette diversité dans la typologie, la nature et l'ordre de grandeur des incertitudes, il n'existe actuellement pas de méthode spécifique et commune pour traiter l'incertitude, et ce, à tous les niveaux (Lloyd & Ries, 2007; Reap et al., 2008). Cette absence de méthodologie systématique dans la gestion de l'incertitude cause une faiblesse dans la publication des résultats d'ACV et dans les prises de décisions se basant sur les résultats d'ACV : l'incertitude ne peut actuellement pas être comparée d'une étude ACV à une autre (Weber, 2012).

Un certain nombre de techniques permettant d'étudier la robustesse des résultats ACV, malgré la présence d'incertitudes, sont toutefois appliquées dans les études d'ACV menées. Parmi elles, on citera principalement l'analyse de sensibilité et l'analyse d'incertitudes. L'analyse de sensibilité permet d'évaluer les changements des résultats, et donc des conclusions de l'étude, dus à des changements de données, d'hypothèses ou de modèles utilisés, quelle que soit la phase de l'ACV

considérée. L'analyse de scénarios est un exemple d'analyse de sensibilité où des changements dans les scénarios choisis (distance de transport ou mode de gestion en fin de vie p.ex.) et leur impact sur les résultats finaux sont évalués. L'analyse d'incertitude peut se faire sous différentes formes, que ce soit par des outils statistiques ou du jugement d'experts, et est principalement appliquée à la phase d'ICV (cf. section 2.3.2).

En définitive, la gestion et l'interprétation de l'incertitude dans une étude ACV se fait selon un certain nombre de techniques différentes où la place de l'expérience et de l'interprétation du praticien est importante, comme le témoigne les quatre moyens identifiés par Finnveden et collègues pour traiter de l'incertitude en ACV (2009) :

- *La manière sociale* : Les différentes parties prenantes trouvent un consensus sur la donnée ou l'hypothèse à adopter;
- *La manière légale* : Même principe que pour la manière sociale à la seule différence que la décision quant aux choix des données ou d'hypothèses revient à la personne représentant l'autorité;
- *La manière scientifique* : Elle consiste à accomplir plus de recherches autour des éléments incertains. L'absence de données et le manque de ressources temporelles et financières sont des limites à cette approche;
- *La manière statistique* : Son objectif n'est pas de s'affranchir de l'incertitude comme pour les trois premières, mais plutôt de la considérer et d'évaluer son impact sur le résultat final.

Bien que les incertitudes émanant de la phase d'ÉICV soient une problématique importante en ACV notamment de par leur ordre de grandeur et qu'elles commencent à être considérées dans certains développements méthodologiques (Bulle, 2012; Manneh, 2011; Roy, 2012), la suite de cet exposé se restreint à la phase d'ICV.

2.3 La gestion de l'incertitude en ICV

2.3.1 Typologies et sources d'incertitude en ICV – le cas particulier de l'incertitude sur les paramètres

Comme mentionné précédemment, l'étape d'ICV présente également les trois typologies principales d'incertitude : l'incertitude sur les modèles, sur les scénarios et sur les paramètres. En plus d'être plus facilement conceptualisable, l'incertitude sur les paramètres – i.e. sur les données – est également celle qui est la plus souvent prise en compte dans les études ACV. C'est la conclusion que Lloyd et Ries (2007) tirent d'un recensement des principales études proposant une méthodologie de prises en compte des incertitudes. Ces incertitudes sur les paramètres proviennent de différents éléments dont :

- l'hétérogénéité des données employées pour mener un ICV, que ce soit en termes de typologie ou de source d'acquisition ;
- le manque d'informations autour de la donnée recherchée (Huijbregts et al., 2001).

En effet, afin de compléter un inventaire, un nombre important de données doit être compilé pour quantifier les échanges réalisés entre les processus élémentaires du système à l'étude et les échanges entre le système et l'environnement. Ces données ont des typologies différentes puisqu'elles peuvent représenter par exemple :

- les émissions à l'environnement d'un processus élémentaire ;
- la consommation énergétique d'un processus élémentaire ;
- la quantité de matière première nécessaire au fonctionnement du processus élémentaire modélisé ;
- les paramètres propres aux activités du processus élémentaire modélisé (le rendement d'une surface agricole, le taux d'utilisation d'un produit p.ex).

Les sources de données d'ICV sont elles aussi nombreuses (cf.

Figure 2-3). Par cette hétérogénéité des sources, la qualité des données utilisées pour mener un ICV n'est pas la même et les causes d'incertitudes diffèrent également. Les différentes caractéristiques influant sur la qualité et l'incertitude associées à une donnée sont recensées à la Figure 2-3.

Que ce soit la qualité ou l'incertitude, elles ne peuvent pas toutes être prises en compte de la même manière puisqu'elles ne s'expriment pas toutes de la même façon. Si les erreurs aléatoires ou systémiques peuvent directement être quantifiables, il n'en est pas de même pour les incertitudes provenant des jugements d'experts, par exemple. Afin de pouvoir prendre en compte « cet ensemble » d'incertitudes et de manque de qualité des données dans la phase d'ICV puis dans la phase d'interprétation d'une ACV, bon nombre de techniques existe, chacune avec leurs propres spécificités.

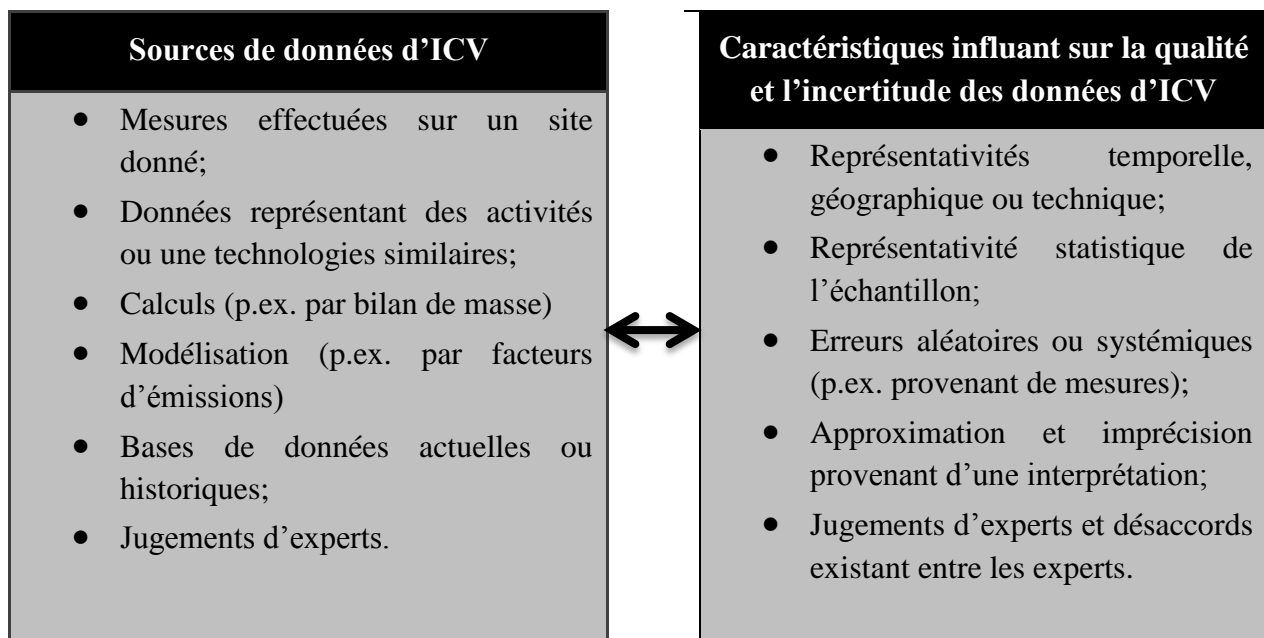


Figure 2-3 Sources de données et caractéristiques influant sur leur qualité et incertitude. Adapté de (Leroy, 2009; Lloyd & Ries, 2007; Maurice et al., 2000; Weidema & Wesnæs, 1996).

2.3.2 Représentation de l'incertitude et de la qualité des données en ICV

Bien qu'il n'existe pas de méthodologie universelle de prise en compte des incertitudes en ACV et plus particulièrement en ICV, les mêmes trois grandes familles que dans le domaine scientifique en général existent pour représenter et traiter l'incertitude et/ou la qualité des données. On trouve donc en ICV l'application d'approches qualitatives, quantitatives et d'approches semi-quantitatives qui se veulent être une combinaison des deux premières.

2.3.2.1 Les approches qualitatives

Le principal objectif des approches qualitatives est de fournir un contrôle de qualité pour les données recueillies et utilisées au cours de la phase d'ICV. Pour la plupart des auteurs, ce contrôle consiste à définir la « carte d'identité » de la qualité de la donnée. Ceci se fait par le biais de critères qui vont permettre d'obtenir des DQI (*Data quality indicator*), i.e. des notes qui sont

associées à la donnée pour chaque critère. Le Tableau 2-2 compare des combinaisons de critères que l'on retrouve dans la littérature.

Si Kennedy et collègues (1996) considèrent un indicateur unique de la qualité de la donnée, les autres experts définissent des critères qui peuvent se recouper suivant les différentes méthodologies. Seuls Rousseaux et collègues (2001) définissent des indicateurs de qualité à différents niveaux (flux, processus, système) qui sont en deuxième temps découpés en « composantes de la qualité ». La norme ISO 14 044 définit, quant à elle, les critères relatifs aux exigences de la qualité des données qui devraient être spécifiés dans la phase de définition des objectifs et du champ de l'étude (par exemple, le critère *incertitude des informations* considère les données mais également les modèles et les hypothèses employées pour réaliser l'étude) (ISO, 2006b).

Dans un premier temps, les DQI vont permettre de qualifier la performance de la qualité par rapport à l'objectif de qualité fixé. Pour ce faire, les experts (qui sont souvent les praticiens qui réalisent le projet) vont noter les données selon chacun des critères identifiés. Les échelles de notations peuvent changer selon la méthodologie employée. Afin de conclure, quant à la qualité d'une donnée, ces notes seront agrégées ou non. Dans la méthodologie développée par Junnila et Horvarth (2003), les notes sont agrégées suivant la phase du cycle de vie. En revanche, les travaux de Wang et collègues (2012) proposent une méthode de pondération des critères selon leur importance qui peut varier selon les domaines dans lesquels ces derniers sont employés. Dans d'autres cas, les DQI sont transformés en indicateurs de qualité. Rousseaux et collègues (2001) se servent des DQI pour définir un coefficient de variabilité et un coefficient d'acceptabilité. Si le premier coefficient va permettre de mettre en lien la qualité de l'inventaire à l'objectif de qualité préalablement fixé, le deuxième permet de mesurer les efforts à mettre en œuvre pour améliorer la qualité globale de l'inventaire.

En ce qui concerne l'agrégation des scores de chacun des critères définis, il n'y a actuellement pas de consensus de la part des auteurs. Comme le témoigne le Tableau 2-2, certains auteurs définissent un indicateur final unique qui pourrait représenter une distance à la qualité spécifiée au préalable (J. R. C., 2010b; Grabowski, et al., 2015; Lewandowska et al., 2004; Maurice et al., 2000; Wang et al., 2012). Pour d'autres auteurs, les critères sont indépendants et l'agrégation des scores peut mener à de mauvaises interprétations de l'information sur la qualité (Cooper et Kahn

se demandent par exemple si, dans l'agrégation des scores, un score 2 est effectivement deux fois meilleur qu'un score 4 (2012)). En effet, pour Weidema et Wesnaes (1996), les scores déterminés ne doivent pas être considérés comme des « valeurs de la qualité » et ne peuvent donc pas être agrégés. Finalement, il est à souligner que l'évaluation de la qualité est faite par rapport à des exigences préalablement fixées et qui peuvent varier selon les méthodologies mais également selon les études ACV menées.

Néanmoins, en plus d'être des indicateurs de la qualité des données, les DQI peuvent permettre de transformer un inventaire déterministe en un inventaire stochastique comme montré dans la section portant sur les approches semi-quantitatives (section 2.3.2.3).

2.3.2.2 Les approches quantitatives

Alors que les approches qualitatives sont principalement destinées à représenter la qualité de l'inventaire réalisé, les approches quantitatives sont utilisées afin de modéliser l'incertitude et la variabilité attachées à une donnée. Pour quantifier ces dernières, bon nombre de techniques sont employées : les statistiques classiques qui se basent sur les distributions de probabilité, les analyses par intervalles, les approches plus marginales des ensembles flous et de la théorie des possibilités et enfin l'analyse bayésienne.

Tableau 2-2 Exemples de critères utilisés en ICV pour coder la qualité d'une donnée

	Critères	Échelle de notation	Agrégation
Weidema et Wesnases (1996)	Fiabilité	De 1 à 5	Non
	Complétude	1 : Bonne qualité	
	Corrélation temporelle	5 : Qualité mauvaise, inconnue ou critère non applicable	
	Corrélation géographique		
	Corrélation technologique		
Maurice et al. (2000)	Critères de fiabilité :		Oui, par pondération : par type de critères les scores ont le même poids ; les deux scores sont ensuite agrégés avec une même pondération
	Représentativité statistique	De 1 à 5	
	Âge de la donnée	1 : Bonne qualité	
	Source de la donnée	5 : Mauvaise qualité	
	Critères d'adéquation		
	Représentativité géographique		
Rousseaux et al. (2001)	Représentativité technique		Non
	14 critères définis selon 4 porteurs:		
	Flux – justesse	De 1 à 5	
	Processus – complétude et représentativité	1 : Bonne qualité	
	Système – Répétabilité	5 : Qualité mauvaise, inconnue ou critère non applicable	
	Incertitude		

Critères	Échelle de notation	Agrégation
Junnila et Horvath (2003)	Méthode d'acquisition Indépendance du fournisseur Représentativité Corrélation temporelle Corrélation géographique Corrélation technologique De 1 à 5 1 : Bonne qualité 5 : Qualité mauvaise, inconnue ou critère non applicable	Oui, scores regroupés par phase du cycle de vie
Lewandowska et al. (2004)	Fiabilité de la source Complétude Corrélation temporelle Corrélation géographique Corrélation technologique De 5 à 1 5 : Bonne qualité 1 : Qualité mauvaise, inconnue ou critère non applicable	Oui en un indicateur de distance à la qualité (qui est lui-même noté), tous les critères ont le même poids
ISO 14044:2006 (2006b)	Facteur temporel Géographie Technologie Fidélité Complétude Représentativité Cohérence Reproductibilité Les critères sont présentés comme des exigences de qualité que les données doivent satisfaire	Non applicable

Critères		Échelle de notation	Agrégation
Source de données			
Incertitude de l'information			
ILCD handbook (J. R. C., 2010b)	Représentativité technologique	De 1 à 5	Oui : les critères avec la même pondération sont sommés, à cela est ajouté 4 fois le score le plus faible
	Représentativité géographique	1 : Très bonne qualité	
	Représentativité liée au temps		
	Complétude	5 : Qualité très faible ou inconnue	
	Incertitude/Précision		
	Adéquation méthodologique et cohérence	0 : Non applicable	
Cooper et Kahn (2012)	Fiabilité et reproductibilité		Non
	Complétude	A : la donnée suit les exigences de qualité	
	Portée temporelle	ou	
	Portée géographique		
	Portée technologique	B : la donnée ne suit pas les exigences de qualité	
	Incertitude		
Wang et al. (2012)	Précision		Aggrégation des critères en appliquant l'outil d'aide à la décision AHP
	Représentativité	De 5 à 1	
	Âge de la donnée	5 : Bonne qualité	
	Méthode d'acquisition		

Critères		Échelle de notation	Agrégation
Grabowski et al. (2015)	Indépendance du fournisseur	1 : Qualité mauvaise ou inconnue	Oui : les critères avec la même pondération sont sommés, à cela est ajouté 4 fois le score le plus faible
	Corrélation géographique		
	Corrélation technique		
	Règle d'inclusion		
	Représentativité technique	De 1 à 5	
	Représentativité géographique	1 : Très bonne qualité	
	Représentativité liée à l'âge	5 : Qualité très faible ou inconnue	
	Complétude		
	Précision/Incertitude	0 : Non applicable	

1. Les approches utilisant les statistiques classiques

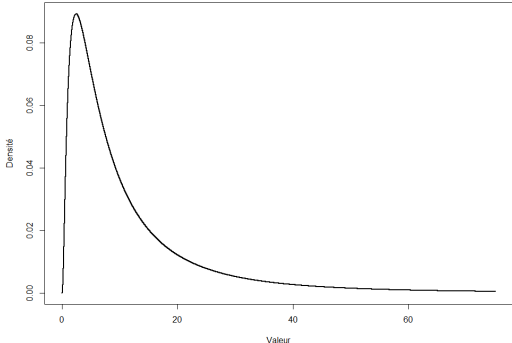
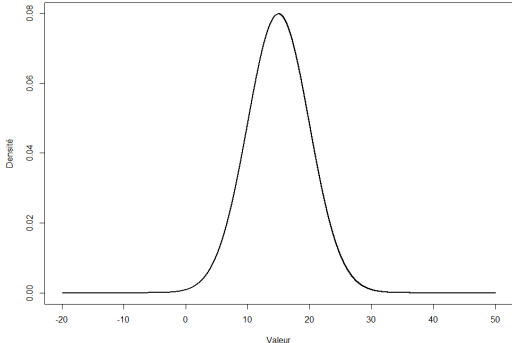
Ces approches sont basées sur la représentation statistique d'une donnée via une fonction de densité de probabilités (PDF). Afin de modéliser la donnée par une PDF et de choisir cette dernière convenablement (notamment par des techniques statistiques permettant de tester la qualité de l'ajustement), la taille de l'échantillon servant à représenter la donnée doit être suffisante pour que l'échantillon soit représentatif. Dans le domaine de l'ACV, si Henriksson et collègues (2014) parlent d'un échantillon devant au minimum contenir 8 points, Weidema et collègues (2003) parlent, eux, de considérer 30 points au minimum; en théorie, l'application des tests statistiques va, quant-à-elle, considérer la taille de l'échantillon. L'approche statistique permet également d'utiliser d'autres composantes des statistiques telles que par exemple :

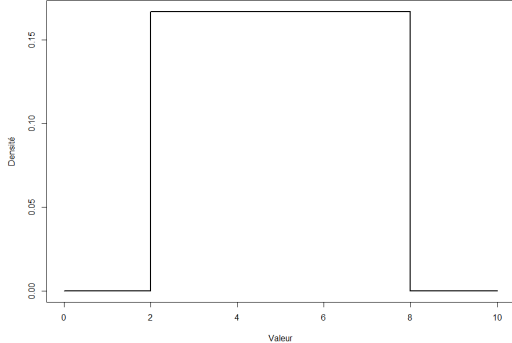
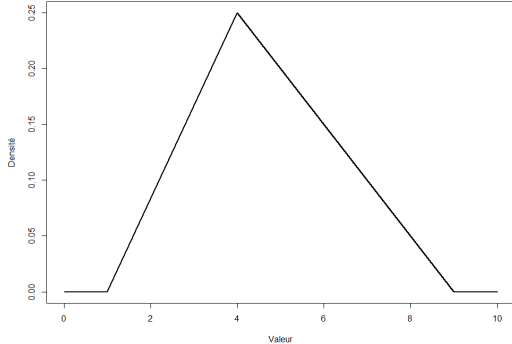
- l'étude de corrélation afin de quantifier les relations de dépendance selon chacun des flux puis de considérer leur importance par rapport à l'incertitude quantifiée issue du modèle complet (Weber, 2012);
- l'analyse en composantes principales, qui une fois l'incertitude modélisée sur le résultat final, permet d'aider à l'interprétation des résultats compte tenu de leur variabilité (Basson & Petrie, 2007);
- l'emploi de méthodes de régression multiple pour déterminer les distributions pour lesquelles peu de données sont disponibles (Venkatesh et al., 2010).

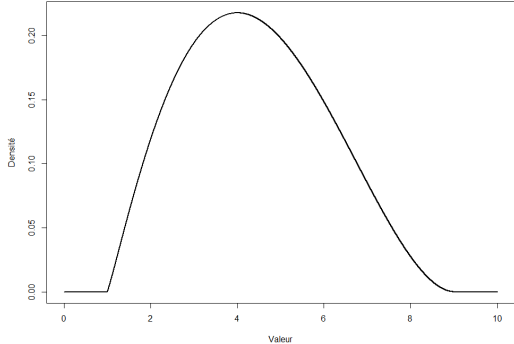
Les quatre distributions souvent utilisées dans le domaine de l'ACV pour modéliser l'incertitude et la variabilité sont la loi log-normale, la loi normale, la loi uniforme et la loi triangulaire (Heijungs & Frischknecht, 2005). D'autres distributions sont également utilisées pour modéliser les données d'inventaire, le Tableau 2-3 fait un recensement de ces distributions utilisées en ICV pour modéliser l'incertitude.

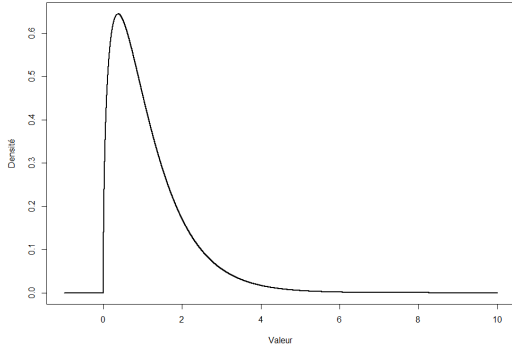
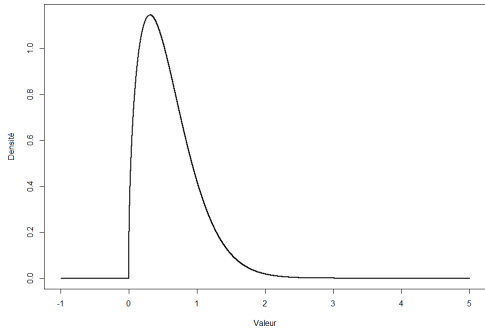
La loi log-normale est largement utilisée en modélisation environnementale puisqu'elle permet de modéliser des paramètres exclusivement positifs. De par son asymétrie et sa capacité à modéliser des valeurs très éparpillées d'échantillons, elle trouve sa place dans des modèles multiplicatifs (Limpert, et al., 2001; MacLeod, et al., 2002; Slob, 1994). Elle est également une des lois les plus utilisées en ICV (cf. Tableau 2-3).

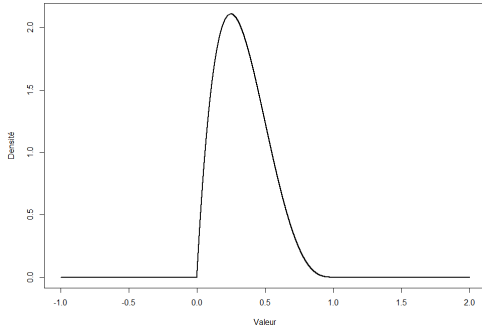
Tableau 2-3 PDF et représentation graphiques de différentes lois utilisées en ICV pour modéliser l'incertitude sur les données

Type de loi et fonction de densité de probabilité	Représentation graphique	Caractéristiques principales	Emploi en ICV
<p>Loi log-normale</p> $f(x, \mu_g, \sigma_g) = \frac{\exp\left(\frac{-(\ln x - \ln \mu_g)^2}{2 \ln^2 \sigma_g}\right)}{\sqrt{2\pi} \ln \sigma_g}$ <p>μ_g : moyenne géométrique σ_g : écart-type géométrique</p>		<p>Non-négative</p> <p>Intervalle non borné : $[0, \infty[$</p> <p>Assymétrie positive</p>	<p>Comme distribution par défaut dans la base de données ecoinvent (Heijungs & Frischknecht, 2005)</p> <p>Pour modéliser les émissions de polluants (Daniels & Higgins, 2002; Hamid, et al., 2013; Sonnemann, et al., 2003)</p> <p>Pour modéliser les consommations énergétiques et les émissions basés sur des tests d'ajustement (Silva, 2011)</p>
<p>Loi normale</p> $f(x, \mu, \sigma) = \frac{\exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$ <p>μ : moyenne arithmétique σ : écart-type arithmétique</p>		<p>Intervalle non borné : $]-\infty, \infty[$</p> <p>Symétrique</p>	<p>PDF alternative à la loi lognormale dans la base de données ecoinvent v3 (Weidema et al., 2013)</p> <p>Pour modéliser les flux économiques (Sonnemann et al., 2003)</p> <p>Pour modéliser les émissions et les données d'activités (Röös, et al., 2010)</p>

Type de loi et fonction de densité de probabilité	Représentation graphique	Caractéristiques principales	Emploi en ICV
<p>Loi uniforme</p> $\begin{cases} f(x, a, b) = \frac{1}{b-a} & \text{pour } a < x < b \\ f(x, a, b) = 0 & \text{sinon} \end{cases}$ <p>a : minimum de la fonction</p> <p>b : maximum de la fonction</p>		<p>Bornée</p> <p>Symétrique</p> <p>À proximité du minimum et du maximum, la probabilité tombe à 0</p>	<p>PDF alternative à la loi lognormale dans la base de données ecoinvent v3 (Weidema et al., 2013)</p> <p>Utilisée quand les paramètres ne sont pas suffisamment connus ou qu'il y a un débat sur la modélisation de leur incertitude (Lloyd & Ries, 2007; Maurice et al., 2000; Mullins, et al., 2010)</p>
<p>Loi triangulaire</p> $\begin{cases} f(x, a, b, c) = \frac{2(x-a)}{(b-a)(c-a)} & \text{pour } a < x < c \\ f(x, a, b, c) = \frac{2(b-x)}{(b-a)(c-b)} & \text{pour } c < x < b \\ f(x, a, b, c) = 0 & \text{sinon} \end{cases}$ <p>a : minimum de la fonction</p> <p>b : maximum de la fonction</p>		<p>Bornée</p> <p>Flexible dans sa forme (dépend du mode)</p> <p>Distribution intuitive qui exagère la probabilité au niveau des bornes</p>	<p>PDF alternative à la loi lognormale dans la base de données ecoinvent v3 (Weidema et al., 2013)</p> <p>Utilisée quand les paramètres ne sont pas suffisamment connus ou qu'il y a un débat sur la modélisation de leur incertitude et que leur mode est connu (Lloyd & Ries, 2007; Maurice et al., 2000; Mullins et al., 2010)</p>

Type de loi et fonction de densité de probabilité	Représentation graphique	Caractéristiques principales	Emploi en ICV
c : valeur la plus probable			
<p>Loi beta PERT</p> $f(x, a, b) = \frac{(x-a)^{\alpha-1} (b-x)^{\beta-1}}{B(\alpha, \beta)(b-a)^{\alpha+\beta-1}}$ <p>a : minimum de la fonction</p> <p>b : maximum de la fonction</p> <p>c : valeur la plus probable</p> <p>et</p> $\alpha = 6 \frac{\mu - a}{b - a} \quad \text{and} \quad \beta = 6 \frac{b - \mu}{b - a}$ $\mu = \frac{a + 4c + b}{6}$		<p>Bornée</p> <p>Flexible dans sa forme (dépend du mode)</p> <p>Moins sensible au niveau des bornes que la distribution triangulaire</p>	<p>PDF alternative à la loi lognormale dans la base de données ecoinvent v3 (Weidema et al., 2013)</p> <p>Utilisée dans un projet industriel pour modéliser l'incertitude sur les paramètres pour des scénarios (Koffler & Kalish, 2012)</p> <p>Utilisée dans une étude de cas lorsque les données avaient une qualité faible (Maurice et al., 2000)</p>

Type de loi et fonction de densité de probabilité	Représentation graphique	Caractéristiques principales	Emploi en ICV
<p>Loi Gamma</p> $f(x, k, \theta) = \frac{x^{k-1} \exp(-x/\theta)}{\Gamma(k)\theta^k}$ <p>k : paramètre de forme θ : paramètre d'échelle</p>		<p>Intervalle non borné : $[0, \infty[$</p> <p>Assymétrie positive</p>	<p>PDF alternative à la loi lognormale dans la base de données ecoinvent v3 (Weidema et al., 2013)</p> <p>Pour modéliser la distribution des particules fines (Hamid et al., 2013)</p> <p>Pour modéliser les émissions des véhicules issues d'échantillons (Lloyd & Ries, 2007)</p>
<p>Loi Weibull</p> $f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & \text{pour } x \geq 0 \\ 0 & \text{pour } x < 0 \end{cases}$ <p>k : paramètre de forme λ : paramètre d'échelle</p>		<p>Intervalle non borné : $[0, \infty[$</p> <p>Assymétrie positive</p>	<p>Pour modéliser la consommation énergétique et les émissions en se basant sur des tests d'ajustement (Silva, 2011)</p>

Type de loi et fonction de densité de probabilité	Représentation graphique	Caractéristiques principales	Emploi en ICV
<p>Loi Beta</p> $f(x, a, b, \alpha, \beta) = \frac{(x - a)^{\alpha-1} (b - x)^{\beta-1}}{B(\alpha, \beta) (b - a)^{\alpha+\beta-1}}$ <p>a : minimum de la fonction</p> <p>b : maximum de la fonction</p> <p>α, β : paramètres de forme</p>		<p>Bornée</p> <p>Large variété de formes dépendamment des paramètres utilisés (symétrique, asymétrie positive ou négative)</p>	<p>Utilisée par défaut dans des méthodologies de prise en compte de l'incertitude due à ses larges variétés de forme (Canter, et al., 2002; Kennedy et al., 1996; Wang & Shen, 2013; Wang et al., 2012)</p>

Ainsi, l'approche statistique permet de représenter la donnée avec son incertitude à l'aide d'une distribution de probabilités. Cette distribution peut être obtenue soit en ayant accès à un échantillon de données, soit en connaissant les paramètres de définition des lois et le comportement général de la donnée. Ces deux manières de définir une fonction de densité de probabilités demandent une collecte de données supplémentaires à réaliser en parallèle de la collecte de données pour l'inventaire ce qui confère à cette approche une limite importante. Certains auteurs proposent même de définir deux distributions de probabilités par donnée : une modélisant la donnée avec ses variabilités temporelles et géographiques et l'autre modélisant la donnée avec son incertitude de mesure, ce qui augmente encore davantage le temps de réalisation de l'étude (Röös et al., 2010). Ce point illustre le fait que la modélisation de la donnée avec son incertitude par le biais de fonctions de densité de probabilité ne permet par contre pas de savoir directement si cette donnée remplit les exigences de qualité pour le système à l'étude.

D'autres auteurs proposent, quant à eux, de modéliser l'ensemble de l'incertitude à l'aide d'une distribution par défaut; tout en se posant la question de l'incertitude supplémentaire qui serait due à ce choix par défaut (Henriksson et al., 2014).

2. L'approche par intervalles

Tableau 2-4 Principes de base d'arithmétique des intervalles

Opération	
Addition	$[a;b] + [c;d] = [a+c ; b+d]$
Soustraction	$[a;b] - [c;d] = [a-d ; b-c]$
Multiplication	$[a;b] \cdot [c;d] = [\min(a \cdot c; a \cdot d; b \cdot c; b \cdot d) ; \max(a \cdot c; a \cdot d; b \cdot c; b \cdot d)]$
Division	$[a;b] / [c;d] = [\min(a/c; a/d; b/c; b/d) ; \max(a/c; a/d; b/c; b/d)]$ $[c;d] \neq [0;0]$

L'approche par intervalles se base sur la représentation de la donnée par ses bornes inférieure et supérieure. Le résultat est une approximation de la donnée sous forme d'un intervalle. Si cette représentation donne une idée de la variabilité, elle ne permet pas de mettre l'accent sur le comportement global et notamment sur la valeur centrale de la donnée ainsi modélisée. Il en est de même pour le résultat final qui va, lui aussi, être exprimé sous forme d'un intervalle. Afin de déterminer ce dernier, l'arithmétique des intervalles est utilisée, ses principes de base sont expliqués au Tableau 2-4 (André & Lopes, 2012; Imbeault-Tétrault, 2010).

3. Les approches utilisant la logique floue et la théorie des possibilités

Ces types d'approches permettent, contrairement aux approches statistiques qui prennent en compte le caractère aléatoire d'une donnée, de représenter le caractère ambigu et flou provenant de la subjectivité associée à la détermination de la donnée. Cette subjectivité provient du jugement d'experts ou encore d'une mauvaise interprétation linguistique des informations associées à la donnée. La théorie des possibilités et des ensembles flous est en mesure de juger et de traiter ce caractère ambigu.

Ainsi la théorie des possibilités permet d'assigner une valeur possible proportionnelle au degré de confiance auquel une sortie pourrait être considérée comme concevable ou non. Cette possibilité se quantifie généralement dans l'intervalle $[0;1]$ ou tout autre intervalle faisant intervenir les bornes inférieure et supérieure de la donnée. Elle représente la crédibilité selon laquelle un événement particulier peut être attendu (Tan, et al., 2007; Tan, Culaba, & Purvis, 2002). André et Lopes (2012) proposent une méthode appliquée à l'ICV pour construire des distributions de possibilités à partir d'informations statistiques sur les données, mais également des moyens de transformer des distributions de possibilités en distributions de probabilités afin de comparer les approches et leurs résultats dans les analyses d'incertitude. Clavreul et collègues (2013) proposent, quant à eux, de combiner les approches probabilistes et possibilistes et de choisir entre l'une ou l'autre dépendamment de la qualité et de la quantité de données présentes dans l'échantillon pour modéliser l'incertitude sur un paramètre.

4. L'approche bayésienne

L'approche bayésienne fait intervenir le théorème de Bayes qui est parfois présenté comme une recette pour apprendre de l'expérience (O'Hagan, 2006). Il peut aussi être caractérisé par le fait qu'il convertit une information *à priori* en une évidence *à posteriori* en se basant sur des données observées (Gill, 2005). Dans sa forme la plus simple, il est exprimé de la sorte :

$$P(E|F) = \frac{P(E)P(F|E)}{P(F)}$$

où E et F représentent deux évènements avec P(E) et P(F) leur probabilité respective,

$P(E|F)$ et $P(F|E)$ sont des probabilités conditionnelles exprimant respectivement la probabilité de E sachant que F est vrai est celle de F sachant que E est vrai.

En utilisant l'écriture intégrale, le théorème de Bayes se note :

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_{\theta} p(y|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}$$

Où θ représente la variable d'intérêt et y une nouvelle information. En pratique, on peut expliquer les paramètres du théorème de la sorte :

- $p(y|\theta)$ est la probabilité de vraisemblance (ou issue du modèle) : elle décrit l'hypothèse selon laquelle les données observées ont été générées en se basant sur le paramètre θ .
- $p(\theta)$ est la probabilité à priori qui décrit la connaissance (souvent subjective) que l'on a sur le paramètre θ .
- $p(\theta|y)$ est la probabilité à posteriori qui modélise l'ensemble des informations que l'on a finalement sur le paramètre θ . La moyenne de la distribution de probabilité peut devenir l'estimé de θ et sa variance ou son intervalle de confiance peut modéliser l'incertitude sur ce paramètre.
- Le dénominateur peut être considéré comme un coefficient de normalisation, il permet de faire en sorte que la probabilité à posteriori appartienne bien à l'intervalle [0,1].

L'utilisation de l'approche bayésienne en ACV pour améliorer notre connaissance sur une donnée et son incertitude est citée par différents auteurs (Björklund, 2002; Huijbregts, 1998; Katz, 2002; Lo, et al., 2005). Les principaux avantages de son utilisation sont qu'elle dépend

directement de l'information à disposition et qu'elle permet de combiner plusieurs types de distributions de probabilités ainsi que différents types d'informations (quantitative, issue du jugement d'experts,...). De plus, grâce aux améliorations dans le domaine de l'analyse numérique par informatique, les temps de calculs ont aujourd'hui largement diminué (Katz, 2002). Néanmoins, très peu d'exemples de son application concrète dans le domaine de l'ACV sont disponibles dans la littérature. Ainsi Ukidwe et collègues et Millers et collègues emploient l'analyse bayésienne pour raffiner leur modélisation. Dans le premier cas, le théorème de Bayes permet de traiter ensemble des informations provenant de différents niveaux (niveau économique, niveau ICV et niveau de l'équipement) (Ukidwe, et al., 2004); tandis que dans le second cas, il est employé pour estimer la probabilité d'adoption d'un scénario donné (Miller, et al., 2013). Lo et collègues ont, quant à eux, employé cette approche sur une étude de cas où, dans un premier temps, la distribution des données avec leur incertitude a été déterminée par jugements d'experts et dans un second temps, pour les données les plus contributrices à l'incertitude, les distributions ont été raffinées grâce à des mesures sur site ou à des rapports de statistiques nationales (Lo et al., 2005).

Le Tableau 2-5 propose une synthèse rapide des différences entre les approches statistiques classiques et les approches bayésiennes.

2.3.2.3 Les approches semi-quantitatives

Il a été vu précédemment que les approches qualitatives servent surtout comme un outil de management de la qualité de la donnée. De leur côté, les approches quantitatives demandent la compilation d'un nombre important de données et ainsi allongent le temps de l'étude qui peut déjà parfois être très important. De plus, ces approches quantitatives ne permettent pas de considérer l'incertitude due à l'utilisation d'une donnée imparfaite dans le contexte de l'étude (c'est le cas par exemple si l'on cherche à modéliser une émission à partir de données d'années antérieures, venant d'une autre aire géographique ou encore à partir d'une technologie semblable mais pas similaire). Les approches semi-quantitatives se veulent une combinaison des deux premières afin de s'affranchir de leurs défauts respectifs.

Tableau 2-5 Différences entre approches statistiques et bayésiennes

	Avantages	Inconvénients
Inférence Bayésienne	Facilité de « mise à jour » Permet de prendre en compte ce que l'on connaît Moins d'hypothèses à considérer pour faire tourner les tests Estimation de paramètres que l'on connaît peu	Temps de calculs importants Introduit la notion de subjectivité Incertitude sur les distributions Incertitude sur les modèles de base (modèles permettant d'obtenir les fonctions de vraisemblance).
Statistiques classiques	Mise en œuvre rapide À préférer lorsque l'information est abondante	Application de nombreux théorèmes et hypothèses Ne permet pas d'introduire de jugement d'experts

Certains experts se basent donc sur les DQI pour identifier les données qui sont les plus importantes contributrices à l'incertitude dans le système étudié. Une fois ces données identifiées, une recherche plus approfondie est réalisée afin de modéliser leur incertitude à l'aide d'une fonction de densité de probabilités (Lewandowska et al., 2004; Maurice et al., 2000).

D'autres auteurs ont développé des techniques permettant de transformer un inventaire basé sur des données obtenues par des valeurs uniques à un inventaire stochastique et ainsi pouvoir appliquer les techniques de statistiques classiques (Kennedy et al., 1996; Weidema & Wesnæs, 1996). Dans les travaux de Kennedy et collègues (1996), si l'incertitude de la donnée est connue, la distribution de probabilités adéquate doit être choisie pour la modéliser. Si ce n'est pas le cas, ils proposent d'utiliser une distribution beta pour laquelle les paramètres de définition sont donnés par rapport au DQI associé à la donnée. Cette méthode est reprise dans les travaux de Wang et collègues (2012) qui transforment leurs DQI pondérés en paramètres de fonction beta pour modéliser la donnée avec son incertitude.

La méthode développée par Weidema et Wesnæs considère les DQI à la fois comme des outils de management de qualité de la donnée et comme des valeurs qui vont permettre de raffiner l'incertitude sur la donnée. En effet, selon leurs travaux, deux types d'incertitudes sont à considérer sur les données utilisées en ICV : l'*incertitude de base* qui représente l'erreur stochastique, i.e. la variabilité inhérente à la donnée et l'*incertitude additionnelle* qui témoigne de l'utilisation d'une donnée imparfaite à la fois pour le contexte de l'étude et par rapport à sa source (non vérifiée, basée sur des estimations...). Les DQI obtenus pour chaque donnée sont les témoins de cette incertitude additionnelle, ils vont être transformés en coefficients de variation (CV) qui, associés à l'incertitude de base – exprimée, elle aussi, en termes de CV – vont permettre de représenter l'incertitude totale de la donnée (Weidema & Wesnæs, 1996).

2.3.2.4 Comparaison des trois approches

Ainsi trois grandes familles de gestion de l'incertitude en ICV ont donc été développées au cours des années, chacune avec leurs avantages et leurs inconvénients comment le témoigne le Tableau 2-6. Parfois, celles-ci sont utilisées en parallèle dans une même étude (notamment les approches quantitatives et semi-quantitatives). Suivant les méthodes, ceci peut être questionnable de par la complétude différente des deux approches (Weber, 2012).

Tableau 2-6 Avantages et inconvénients des différentes méthodes de traitement de l'incertitude en ICV

Approches	Avantages	Inconvénients
Qualitative	Outil de management de la qualité	Ne permet pas une représentation de l'incertitude
	Mise en œuvre rapide	Subjectivité de la notation du praticien
Quantitative	Représentation la plus fidèle de la variabilité inhérente à la donnée	Coûteuse en termes de données et donc de temps et d'énergie
	Permet l'expression quantitative de l'incertitude sur le résultat final	Nécessité pour le praticien de posséder des notions en statistiques et probabilités
Semi-quantitative	Représentation quantitative de l'incertitude	Ne permet pas de considérer l'incertitude due à l'utilisation d'une donnée imparfaite
	Permet l'expression quantitative de l'incertitude sur le résultat final	Subjectivité de la notation du praticien
Semi-quantitative	Mise en œuvre rapide	Nécessité pour le praticien de posséder des notions en statistiques et probabilités
	Permet d'intégrer de façon quantitative les aspects de qualité des données	Complétude différente selon les données

2.3.3 Méthodes de propagation des incertitudes sur les paramètres

Une fois les incertitudes identifiées et modélisées, il s'agit de voir leur effet sur le résultat final, et ce, en prenant en compte leur propagation et donc leur interaction dans le système étudié puisque, comme le soulignent Citroth et collègues (2004), deux effets inverses peuvent être observés:

- Les incertitudes peuvent s'amplifier en s'ajoutant tout au long du modèle ;
- L'incertitude peut diminuer en ajoutant des données dont l'incertitude varie en sens contraire.

La méthode la plus utilisée pour étudier la propagation des incertitudes et l'incertitude sur le résultat final (que ce soit dans le domaine de l'ACV ou celui de l'analyse de risque) est l'analyse de type Monte-Carlo. Elle est implémentée dans deux des logiciels les plus utilisés par les praticiens de l'ACV à savoir SimaPro³ et GaBi⁴. La méthode de simulation de Monte Carlo prend en compte l'incertitude sur les paramètres d'entrée et les relations qui les lient afin d'exprimer l'incertitude sur les paramètres de sortie du système à l'étude en termes de distributions de probabilités. Pour ce faire, les données d'entrée sont modélisées avec leur incertitude par des distributions de probabilités. À chaque pas de la simulation de Monte Carlo, les valeurs des données d'entrée sont obtenues par tirage aléatoire selon cette distribution de probabilité. À chaque pas une valeur de sortie est donc calculée. À la fin de la simulation, une distribution de probabilités correspondant à ces valeurs peut être tracée, permettant ainsi de représenter la valeur de sortie avec son incertitude. Malgré le fait que cette approche est une approche complète, afin de tendre vers un résultat stable, un nombre important de pas dans cette simulation doit être effectué ce qui demande des ressources en temps et en capacité informatique non négligeables.

Des approches analytiques se basant sur des développements de Taylor du premier ordre ont également été développées pour une application en ACV. Ces approches demandent moins de ressources en temps et en capacité informatique que les simulations de Monte Carlo. Ce développement de Taylor (souvent du premier ordre) permet à cette approche de se baser sur la propagation de la variance par le biais de l'écart-type géométrique. L'incertitude sur le résultat final est, elle aussi, exprimée en termes d'écart-type géométrique. Ces méthodes permettent

³ <http://www.pre-sustainability.com/simapro-lca-software>

⁴ <http://www.gabi-software.com>

également de faire des analyses de sensibilité sur les paramètres et d'identifier les paramètres d'entrée qui contribuent le plus à l'incertitude sur les paramètres de sortie. Néanmoins, comme le développement de Taylor effectué est un développement du premier ordre, l'approche est plus adéquate pour des valeurs d'incertitude faibles sur les paramètres d'entrée. (Hong et al., 2010; Imbeault-Tétreault, et al., 2013; MacLeod et al., 2002; Slob, 1994).

Heijungs and Lenzen (2014) ont récemment comparé l'utilisation d'approches numériques (de type Monte Carlo) et d'approches analytiques pour la propagation de l'incertitude en ACV. S'ils remettent en cause l'hypothèse de log-normalité citée dans les travaux touchant à l'approche analytique, leurs principales conclusions est que les deux approches doivent être incorporées dans les logiciels ACV puisque, même si l'analyse numérique donne accès à plus d'informations sur le résultat, elle demande plus d'informations sur les données d'entrée tandis que les méthodes analytiques sont plus rapides à mettre en œuvre et demandent moins d'informations.

Ciroth et collègues (2004) ont, quant à eux, proposé une approche plus générale qui se base non plus sur une unique distribution (la log-normale ici), mais sur les valeurs des CV. L'approche est basée sur des approches analytiques du premier et du deuxième ordre et sur la simulation de Monte Carlo pour analyser la propagation de l'incertitude. Le choix de la méthode à utiliser se base sur des valeurs seuils qui dépendent des CV calculés à chaque étape.

Les techniques de propagation de l'incertitude sur les entrants pour évaluer l'incertitude sur les résultats finaux sont ainsi de diverses natures. Qu'elles soient basées sur la simulation numérique ou sur des approximations du premier ou second ordre, elles demandent au minimum que les paramètres statistiques des données d'entrée soient connus (écart-type géométrique ou CV). Elles ne peuvent donc être utilisées qu'avec certaines des méthodes de gestion de l'incertitude citées à la section 2.3.2

2.3.4 La gestion de l'incertitude dans les bases de données d'inventaire

Les bases de données d'inventaire sont la principale source de données utilisée par les praticiens pour modéliser leur système à l'étude et réaliser l'inventaire des émissions de ce dernier. Ces bases de données comportent des processus élémentaires spécifiques ou génériques, agrégés ou non entre eux et modélisés avec l'ensemble des flux élémentaires et économiques qu'ils soient

entrants ou sortants. Une large partie de l'inventaire est donc réalisée à partir de ces bases de données qui ne considèrent pas toutes l'incertitude de la même manière.

Le guide émis par l'Union Européenne (*International Reference Life Cycle Data System (ILCD) Handbook - Specific guide for Life Cycle Inventory datasets*) pour l'élaboration de processus élémentaires dans le cadre de base de données spécifie que la qualité de chaque donnée doit être évaluée afin de pouvoir définir la qualité globale d'un processus élémentaire. En termes d'incertitude, il est stipulé que la distribution de probabilités modélisant une donnée avec son incertitude doit être définie si possible, i.e. si l'information disponible le permet. Une méthode semi-quantitative est également proposée afin de transformer les notations de qualité en écart-type (European Commission, 2010). La base de données européenne (ELCD) se base en partie sur les recommandations de ce guide. Néanmoins, si des informations sur la qualité des données sont disponibles ce n'est pas le cas pour l'incertitude puisqu'aucun champ la concernant n'est présent une fois les processus téléchargés. En ce qui concerne la base de données américaine (*US Life cycle inventory database*), l'incertitude n'est définie pour une donnée que si celle-ci a été quantifiée. Dans ce cas, quatre distributions de probabilités peuvent être utilisées pour modéliser l'incertitude (National Renewable Energy Laboratory, 2004). La base de données japonaise ne semble, quant à elle, pas prendre en compte l'incertitude sur les données d'inventaire mis à part via un champ minimum/maximum qui permet d'évaluer l'intervalle dans lequel se situe la valeur (Narita, et al., 2004; Sugiyama, et al., 2005). En termes de bases de données spécifiques à un secteur industriel particulier, citons la base de données *Agri-footprint* spécifique au domaine de l'agriculture. Là aussi, la modélisation quantitative de l'incertitude ne se fait pas de manière systématique. Ainsi, seulement lorsqu'elle est disponible (quantitativement ou par jugement d'experts), l'incertitude sur les modèles et les paramètres est prise en compte (Blonk, 2011).

En ce qui concerne les deux bases de données les plus largement utilisées (GaBi et ecoinvent), leur traitement de l'incertitude sur les données diffère également. Pour GaBi, la qualité doit être notée selon trois critères (précision, complétude et représentativité) et l'incertitude n'est à représenter que si son évaluation quantitative est possible (GaBi, 2007). Pour ecoinvent en revanche, le traitement de l'incertitude est basé sur l'approche semi-quantitative pedigree inspirée du système NUSAP et adaptée pour l'ACV par Weidema et Wesnæs (1996). Ce système permet une évaluation quantitative de l'incertitude pour l'ensemble des entrants (Frischknecht et al., 2005).

La base de données ecoinvent et ses initiatives régionales (Australie, Québec,...) est donc la seule qui permette l'évaluation quantitative de l'incertitude sur l'ensemble des paramètres des flux élémentaires et économiques, à la fois pour les données d'arrière-plan et pour les données d'avant-plan. En effet, pour les autres bases de données, le renseignement quantitatif de l'incertitude n'est pas obligatoire. Néanmoins, ces bases de données, de par les logiciels dans lesquels elles sont implémentées, permettent de conduire des simulations de Monte Carlo, bien qu'une seule partie de l'information sur l'incertitude soit modélisée, entraînant ainsi de potentielles mauvaises interprétations des résultats puisque seule l'information sur l'incertitude effectivement quantifiée est traitée. Le fait qu'ecoinvent soit la seule base de données qui permette cette évaluation exhaustive et systématique de l'incertitude s'explique par la nature de son système de traitement de l'incertitude.

2.4 Une approche de gestion semi-quantitative de l'incertitude appliquée à la base de données ecoinvent : l'approche pedigree

2.4.1 Le système NUSAP

L'approche pedigree est issue d'une méthodologie plus générale traitant de la qualité et de l'incertitude de manière quantitative et qualitative : le système NUSAP développé par Funtowicz et Ravetz (1990). Ce système se veut un système de notation mathématique mis en place pour pallier les lacunes du système conventionnel : un chiffre en lui-même ne peut représenter la réalité d'une donnée puisqu'il ne tient pas compte de la variabilité par exemple de cette dernière, et de témoigner de la complexité des systèmes desquels émane la donnée étudiée. L'objectif de ce système est à la fois de compléter les connaissances des analyses quantitatives de l'incertitude en réalisant une analyse qualitative, mais aussi de mieux communiquer les informations liées à la qualité et à l'incertitude, et ce, dans un langage compréhensible par l'ensemble des décideurs. Pour faire ce diagnostic complet de la qualité et de l'incertitude, ce système se base sur les cinq lettres de son acronyme (*Numeral Unit Spread Assessment Pedigree*) qui sont chacune rattachées à un élément différent de la donnée. Si le N et le U sont à associer à la valeur numérique, déterministe de la donnée ainsi qu'à son unité, le S est à relier à variance et plus spécifiquement à l'inexactitude statistique, i.e. la variance issue de l'échantillon. Ainsi l'incertitude modélisée par

le S est une incertitude située à un niveau technique (d'expérimentations), alors que le A et le P témoignent de l'incertitude épistémique et méthodologique puisqu'ils sont à relier aux jugements qualitatifs de l'expert (A) et à son évaluation (P) qui se fait via une matrice pedigree. Cette modélisation des jugements qualitatifs ne peut se faire de manière automatique et nécessite la réflexion de l'expert (Funtowicz & Ravetz, 1990; van der Sluijs, et al., 2004; van der Sluijs et al., 2005).

Concrètement, l'évaluation du jugement de l'expert et de sa connaissance de la donnée étudiée se fait selon une matrice pedigree où les critères d'évaluation (indicateurs) se retrouvent en ligne et les notes en colonne (scores). Les cellules de la matrice contiennent une description de la qualité attendue pour chaque score et par indicateur. Suivant les domaines dans lequel le système NUSAP et la matrice pedigree sont employés pour évaluer la qualité et l'incertitude, les indicateurs ainsi que leur notation peuvent différer. Le site <http://nusap.net/> donne des exemples de matrice pedigree utilisées dans le domaine de la recherche, de l'évaluation environnementale ou encore de l'agroalimentaire. Dans le domaine de l'ACV, une matrice pedigree pour l'évaluation des données d'inventaire a été développée dans les années 90 (Weidema & Wesnæs, 1996). Citroth (2009) s'en est inspiré pour définir une matrice pedigree applicable dans le domaine de l'analyse des coûts du cycle de vie, le penchant économique de l'ACV.

2.4.2 Application de l'approche pedigree pour la gestion de l'incertitude dans la base de données ecoinvent

L'approche semi-quantitative utilisée pour représenter l'incertitude sur un paramètre dans la base de données ecoinvent est basée sur l'utilisation de fonctions de densité de probabilités et sur celle de la matrice pedigree développée par Weidema et Wesnæs. Elle permet de considérer les deux types d'incertitudes définies par Weidema et Wesnæs, à savoir : l'*incertitude de base* et l'*incertitude additionnelle* (cf. section 2.3.2.3). L'incertitude additionnelle est codée selon cinq indicateurs indépendants confinés dans une matrice pedigree représentée au Tableau 2-7. Les cinq indicateurs représentent les champs suivants :

- La justesse, précision met l'accent sur la manière dont la donnée a été obtenue et si des vérifications ont été effectuées ou non;

- La complétude témoigne de la qualité statistique de la donnée, i.e. de la représentativité de l'échantillon disponible;
- La corrélation temporelle lie l'année durant laquelle l'activité a eu lieu dans le système de produit à l'année durant laquelle la donnée a été obtenue;
- La corrélation géographique lie l'aire géographique de l'activité à l'aire pour laquelle la donnée est obtenue, et ce, d'un point de vue législatif notamment;
- La corrélation technologique lie la technologie pertinente à l'étude à la technologie représentée par la donnée.

En d'autres termes, les indicateurs de corrélation lient la donnée à la qualité aux objectifs de qualité de l'étude alors que les indicateurs de justesse et complétude sont à lier à la qualité de l'échantillon. La notation de ces indicateurs se fait sur une échelle de 1 à 5, 1 étant le meilleur score.

Tableau 2-7 Matrice pedigree, traduite, appliquée dans la base de données ecoinvent et issue des travaux de Weidema et Wesnaes (1996)

Score	1	2	3	4	5 (par défaut)
Justesse, précision	Donnée vérifiée et basée sur des mesures	Donnée vérifiée partiellement et basée sur des hypothèses OU données non vérifiées et basées sur des mesures	Donnée vérifiée partiellement basée sur des estimations qualifiées	Estimation qualifiée (par un expert industriel par exemple)	Estimation non qualifiée
Complétude, exhaustivité	Donnée représentative de tous les sites liés au marché considéré sur une période adéquate afin	Donnée représentative de plus de 50% des sites liés au marché considéré sur une période	Donnée représentative de quelques sites liés au marché considéré (<50%) ou	Donnée représentative d'un seul site relié au marché considéré OU de quelques sites mais pour	Représentativité inconnue ou donnée d'un petit nombre de sites ET de périodes plus

	de réduire les fluctuations	adéquate afin d'éviter les fluctuations	représentative de plus de 50% des sites mais pour des périodes plus courtes	des périodes plus courtes	courtes
Corrélation temporelle	Donnée avec moins de trois ans de différence avec l'année de l'étude	Donnée avec moins de six ans de différence avec l'année d'étude	Donnée avec moins de dix ans de différence avec l'année d'étude	Donnée avec moins de quinze ans de différence avec l'année d'étude	Âge de la donnée inconnu OU donnée avec plus de quinze ans de différence avec l'année d'étude
Corrélation géographique	Donnée issue de la zone d'étude	Donnée moyenne d'une zone plus large dans laquelle la zone à d'étude est incluse	Donnée issue d'une zone aux conditions de production similaires	Donnée issue d'une zone aux conditions de production légèrement similaires	Donnée issue d'une zone inconnue OU d'une zone avec des différences importantes
Corrélation technologique	Donnée issue d'entreprises, de processus et de matériels à l'étude	Données issues de processus et de matériels à l'étude (même technologie) mais d'entreprises différentes	Donnée issue de processus et de matériels à l'étude mais de technologies différentes	Donnée associée à des processus et des matériels connexes	Donnée reliée à des processus à l'échelle du laboratoire ou à des technologies différentes

Concrètement, l'application de l'approche pedigree dans la base de données ecoinvent v2 se fait selon la disponibilité ou non de données statistiques pour le paramètre à l'étude et si ces données statistiques sont issues ou non du contexte de l'étude, on distingue ainsi trois cas de figures différents :

1. Les informations statistiques sont disponibles et sont issues du contexte de l'étude.

Dans ce cas, la donnée peut être modélisée, à partir de l'échantillon disponible, avec son incertitude par une distribution de probabilités choisies parmi les distributions suivantes : log-normale, normale, uniforme, triangulaire, beta PERT ou gamma. Les scores pedigree sont égaux à 1 pour chacun des indicateurs étant donné que les points obtenus reflètent parfaitement la situation à l'étude.

2. Les informations statistiques sont disponibles mais ne correspondent pas au contexte de l'étude.

De la même manière que précédemment, dans un premier temps la donnée est modélisée avec son incertitude par une distribution de probabilités puis la matrice pedigree est remplie afin de coder la qualité de la donnée par rapport à une donnée idéale.

3. Les informations statistiques ne sont pas disponibles.

Tout comme précédemment une distribution de probabilités doit dans un premier temps être choisie en se basant sur la littérature ou les caractéristiques de la donnée. Si la variabilité, i.e. l'incertitude de base n'est pas connue, ecoinvent propose une série de facteurs par défaut par type de flux et type de processus, cf. Tableau 2-8. En effet, la variabilité du paramètre quantitatif pour un flux donné dépend de la nature de ce dernier. Par exemple, le facteur d'incertitude de base associé à l'émission de métaux lourds dans l'eau est plus important que celui représentant l'émission du CO₂ à l'air étant donné que les caractéristiques de cette dernière sont connues de manière plus précise. Enfin, si la donnée ne provient pas du contexte de l'étude, la matrice pedigree doit être renseignée.

Tableau 2-8 Tableau des facteurs d'incertitude de base appliqués par défaut dans la base de données ecoinvent (Weidema, et al. 2010)

Classe d'entrants ou de sortants	Processus de combustion	Processus	Processus agricoles
Demande en :			
Énergie thermique, électricité, produits semi-finis, matériel, traitement de déchets	1.05	1.05	1.05
Services de transport (en tkm)	2.00	2.00	2.00
Infrastructure	3.00	3.00	3.00
Ressources :			
Source d'énergie primaire, métaux, sels	1.05	1.05	1.05
Utilisation des terres, occupation	1.50	1.50	1.10
Utilisation des terres, transformation	2.00	2.00	1.20
Polluants émis à l'eau :			
BOD, COD, DOC, TOC, composés inorganiques		1.50	
Hydrocarbures, hydrocarbures aromatiques polycycliques		3.00	
Métaux lourds		5.00	1.80
Pesticides			1.50
Nitrate, Phosphate			1.50
Polluants émis au sol :			
Huile, hydrocarbures		1.50	
Métaux lourds		1.50	1.50
Pesticides			1.20
Polluants émis à l'air :			
Dioxyde de carbone	1.05	1.05	
Dioxyde de soufre	1.05		

Classe d'entrants ou de sortants	Processus de combustion	Processus agricoles
Composés organiques volatiles non-méthaniques	1.50	
Monoxyde/ dioxyde d'azote, protoxyde d'azote	1.50	1.40
Méthane, ammoniac	1.50	1.20
Hydrocarbures	1.50	2.00
Particules fines (>10 µm)	1.50	1.50
Particules fines (< 10 µm)	2.00	2.00
Particules fines (<2.5 µm)	3.00	3.00
Hydrocarbures aromatiques polycycliques	3.00	
Monoxyde de carbone, métaux lourds	5.00	
Émissions inorganiques		1.50
Radionucléides		3.00

En plus d'être des indicateurs de qualité de la donnée, les scores pedigree vont pouvoir être convertis en facteurs d'incertitude additionnelle qui, associés à l'incertitude de base, vont exprimer l'incertitude totale sur la donnée. Actuellement cette association incertitude de base / incertitude additionnelle ne se fait que pour des données distribuées de manière log-normale – loi utilisée pour représenter près de 70% des flux et de leur incertitude dans ecoinvent v2 (Ecoinvent, 2010). Cette compilation se base sur la propagation des écarts-type géométriques pour une loi log-normale. Pour ce faire, les facteurs d'incertitude de base présentés au Tableau 2-8 sont exprimés en carré d'écart-type géométrique. Les scores pedigree sont également convertis en carré d'écarts-type géométriques – les facteurs d'incertitude additionnelle - (cf. Tableau 2-9). L'ensemble de ces facteurs d'incertitude proviennent principalement de jugements d'experts.

L'Équation 2-1 permet, quant à elle, d'exprimer l'incertitude totale UF_T par rapport au facteur d'incertitude de base UF_b et aux facteurs d'incertitude additionnelle UF_i . Si l'on considère une

donnée parfaite (i.e. avec l'ensemble de ces scores pedigree égaux à 1), l'incertitude totale est égale à l'incertitude de base.

Tableau 2-9 Tableau de conversion des scores pedigree en facteurs d'incertitude additionnelle

Score	1	2	3	4	5
Justesse	1.00	1.05	1.10	1.20	1.50
Complétude	1.00	1.02	1.05	1.10	1.20
Corrélation temporelle	1.00	1.03	1.10	1.20	1.50
Corrélation géographique	1.00	1.01	1.02	1.05	1.10
Corrélation technologique	1.00	1.05	1.20	1.50	2.00

$$UF_T = \exp \left(\sqrt{\ln^2 UF_b + \sum_{i=1}^n \ln^2 UF_i} \right) \quad \text{Équation 2-1}$$

Dans sa version 3, la base de données ecoinvent a choisi d'exprimer les facteurs d'incertitudes en termes de variance plutôt qu'en termes de carré d'écarts-type géométriques. Si cela ne change rien à l'approche en tant que telle, les facteurs d'incertitudes ont dû être traduits en termes de variance et l'Équation 2-1 a dû être adaptée (Weidema et al., 2013).

Bien que critiquée par certains auteurs notamment par le fait que l'ACV ne suit pas la mise en application du schéma NUSAP (cf. section 2.4.1) et du coup en ferait une mauvaise application par l'utilisation d'une matrice pedigree (Henriksson et al., 2014), cette approche pedigree peut être à la fois utilisée comme un outil de gestion de la qualité de la donnée et comme manière de transformer l'inventaire en un inventaire stochastique. Elle permet ainsi de modéliser l'incertitude sur chacun des flux élémentaires et économiques; de cette manière, l'ensemble de ces derniers peuvent être considérés dans les méthodes de propagation de l'incertitude et ainsi influencer l'incertitude sur les résultats.

2.5 Problématiques issues de la revue de littérature

2.5.1 La nécessité d'une approche permettant la gestion de l'incertitude sur l'ensemble des données utilisées en ICV

Comme souligné dans la revue de littérature, une multitude d'approches de gestion de l'incertitude sur les données d'ICV existent actuellement, elles sont diverses et ont chacune leurs avantages et leurs inconvénients. Les approches les plus complètes, permettant de quantifier l'incertitude sur l'ensemble des données, sont très coûteuses en termes de compilation de données et de temps et ne permettent pas de quantifier l'incertitude associée à l'utilisation d'une donnée imparfaite. Certaines approches de traitement ne la considèrent que sur des données pour lesquelles l'incertitude a pu être quantifiée et mènent les analyses d'incertitudes en ne prenant en considération que ces données, ce qui amène une limite sur la représentativité de l'incertitude du résultat final. Or, il a été vu qu'il était important de considérer l'incertitude sur le résultat final pour la prise de décisions, et ce, de la manière la plus juste et la plus fidèle possible. En effet, en considérant l'incertitude, la décision pourra se faire avec un degré de confiance inversement proportionnel à l'incertitude sur les résultats et la crédibilité de la décision en sera ainsi accentuée.

Si les approches quantitatives sont idéales en théorie pour obtenir des résultats avec une évaluation de l'incertitude plus précise, en pratique elles sont impossibles à mettre en œuvre sur l'ensemble des données à considérer pour réaliser un ICV. Les approches qualitatives permettent seulement d'avoir une image de la qualité globale de l'inventaire et donc des résultats exprimés en fonction des objectifs de qualité définis dans la phase de définition des objectifs et du champ de l'étude. Les qualités globales définies et obtenues sont donc propres à une étude et peuvent varier dépendamment des objectifs fixés. Les approches semi-quantitatives semblent donc être le meilleur compromis pour transformer l'inventaire en un inventaire stochastique et modéliser l'incertitude sur l'ensemble des données d'inventaire à l'aide de distributions de probabilités ; que des informations concernant l'incertitude soient disponibles ou non. De plus, ces approches permettent de considérer la représentativité de la donnée utilisée dans le contexte de l'étude.

Il n'existe actuellement pas d'approche universelle permettant de considérer l'incertitude sur les données d'inventaire (Williams et al., 2009). Si certains ont proposés des cadres pour la gestion de l'incertitude (Huijbregts et al., 2001; Huijbregts, et al., 2003), chacun y va de sa méthodologie pour modéliser l'incertitude et sa propagation jusqu'au résultat final. Une approche néanmoins largement utilisée, de par sa présence dans une base de données d'inventaire reconnue et utilisée par les praticiens de l'ACV, existe. Cette approche nommée *pedigree* est semi-quantitative et mise en œuvre dans les analyses d'incertitude réalisées à partir des inventaires de la base de données ecoinvent. Cette base de données est largement utilisée dans les principaux logiciels d'ACV. Elle est transparente et possède à la fois des données globales et régionales. Néanmoins, telle qu'appliquée dans la version 2 de la base de données, cette approche *pedigree* possède un certain nombre de limites.

2.5.2 Les limites actuelles de l'approche *pedigree*

Telle que décrite dans la section 2.4, l'approche *pedigree* possède un certain nombre de limites :

- Elle n'est actuellement valable que pour des distributions log-normales. Bien que la distribution log-normale possède de nombreux avantages pour la modélisation de paramètres physiques, les autres distributions (cf. Tableau 2-3) peuvent être plus appropriées si les informations pour modéliser l'incertitude sont disponibles ;
- Les différents facteurs d'incertitude appliqués dans la « méthodologie *pedigree* » (facteurs d'incertitude de base et facteurs de conversion d'incertitude additionnelle) ne possèdent pas d'assise empirique solide. En effet, ils ont été obtenus par jugement d'experts. Et bien que les avis d'experts soient un moyen d'obtention d'information valide, ces avis n'ont pas été vérifiés par des données empiriques ;
- Même si dans le cadre d'un exemple d'application, une matrice *pedigree* spécifique à un secteur industriel est citée dans les travaux de Weidema (1998) ; la matrice *pedigree* est actuellement définie de manière unique quel que soit le type de flux considéré. Si pour l'indicateur de complétude cela ne pose pas de problème, il pourrait ne pas en être de même pour les autres. En effet, en considérant par exemple l'indicateur de corrélation temporelle, les descriptions des cellules de la matrice quantifient l'âge de la donnée (3, 6,

10, 15 ans et plus). Or l'incertitude due à l'âge de la donnée utilisée pour représenter la donnée parfaite peut dépendre de la nature de cette dernière. La variabilité temporelle d'une quantité de bois sur trois ans pourrait différer de celle d'une quantité de nano-matériel pour cette même période ;

- L'approche n'est actuellement valable que pour les données représentant les flux. D'autres données quantitatives sont néanmoins nécessaires pour modéliser un inventaire du cycle de vie, il s'agit, par exemple, des paramètres permettant de relier l'unité fonctionnelle aux valeurs de ces flux.

Malgré l'ensemble de ces limites dans l'approche pedigree telle qu'appliquée actuellement, cette approche a le principal avantage de ne discriminer aucune donnée liée à l'inventaire pour le calcul de l'incertitude sur le résultat final. De plus, elle permet de considérer l'ensemble des caractéristiques influant sur l'incertitude, la qualité et la représentativité des données utilisées en ICV. Ce projet vise donc à repenser l'approche pedigree et son application, notamment dans les bases de données d'inventaire, afin de s'affranchir de ses limites et quantifier l'incertitude sur les données entrantes du système de manière plus précise.

CHAPITRE 3 OBJECTIFS DU PROJET ET MÉTHODOLOGIE GÉNÉRALE

3.1 La première phase du « projet pedigree »

Ce projet de doctorat fait suite à la première phase d'un projet commandité (appelé ici « projet pedigree ») en 2010 par le *ecoinvent centre* qui est l'organisme en charge de la gestion de la base de données ecoinvent. Le projet pedigree avait pour mandat de légitimer la structure de gestion de l'incertitude. Au cours de sa première phase, une équipe s'est penchée sur le développement de facteurs d'incertitude additionnelle empiriques (Ciroth, et al., 2013). Si les travaux ont permis de proposer de nouveaux facteurs d'incertitude additionnelle, certaines limites demeurent :

- Les facteurs d'incertitude obtenus restent des facteurs génériques, aucune distinction n'a été faite par secteurs d'activités;
- Peu de bases de données ont été étudiées pour l'ensemble des indicateurs (une à trois par indicateurs), de plus ces bases de données étaient souvent spécifiques à un secteur d'activités. Ceci peut entraîner un problème de représentativité des facteurs génériques obtenus;
- Seuls les facteurs d'incertitude additionnelle ont été déterminés, le cas de l'incertitude de base n'a pas été traité dans ces premiers travaux.

Suite à cette première étude partielle, il s'agit d'aller plus loin afin de raffiner l'approche pedigree telle qu'appliquée dans la base de données ecoinvent.

3.2 Définition des objectifs et du projet de recherche

Ce projet s'inscrit donc dans une volonté d'améliorer la confiance des praticiens et des décideurs envers les résultats d'ACV par une quantification plus juste de l'incertitude. De manière plus spécifique, l'objectif global de ce projet est de développer et légitimer par une assise scientifique la structure de gestion semi-quantitative de l'incertitude sur les données d'inventaire qu'est l'approche pedigree et notamment pour son application dans les bases de données d'inventaire. Cet objectif global se base sur les trois questions de recherche suivantes :

1. Quel est l'impact du choix de la distribution log-normale comme distribution par défaut pour modéliser une donnée avec son incertitude
2. Comment l'approche pedigree peut-elle être appliquée à d'autres distributions de probabilités?
3. Comment les informations concernant les incertitudes des données d'inventaire peuvent-elles être améliorées en se basant à la fois sur une assise empirique et sur les jugements d'experts connus?

Afin de répondre à ces questions, les objectifs spécifiques suivants sont définis et seront traités dans la suite de ce manuscrit :

1. Légitimer l'usage de la distribution log-normale comme distribution par défaut pour modéliser une donnée d'inventaire avec son incertitude;
2. Déterminer comment l'approche pedigree telle qu'appliquée actuellement dans la base de donnéesecoinvent peut être appliquée à des distributions de probabilités autre que log-normale;
3. Développer une méthode d'analyse de données permettant de déterminer, avec une assise empirique, et de mettre à jour les facteurs d'incertitudes, et ce par secteur d'activités.

Le développement méthodologique ainsi que les résultats obtenus pour chacun de ces objectifs spécifiques sont présentés sous forme d'articles scientifiques aux chapitres suivants. La section suivante présente, quant à elle, la méthodologie générale proposée afin de répondre aux questions de recherche soulevées et aux objectifs définis.

3.3 Méthodologie générale

Afin de répondre aux questions de recherche et aux objectifs spécifiques de ce projet, il s'agit d'élaborer des développements méthodologiques qui permettent *in fine* de mettre à la disposition de l'utilisateur un outil – l'approche pedigree – robuste de quantification systématique de l'incertitude sur les données d'inventaire. Telle que présentée à la section 2.4, l'approche pedigree peut se décomposer en trois étapes, possédant chacune des limites. La réponse à l'objectif de recherche passe donc par le traitement et l'amélioration de ces trois étapes en

parallèle, tel que décrit à la Figure 3-1. La Figure 3-2 présente, quant à elle, le déroulement méthodologique à appliquer pour répondre à l'objectif général et aux questions de recherche citées précédemment. Les développements méthodologiques mis en œuvre pour répondre à ces trois objectifs spécifiques peuvent être traités de manière indépendante. Néanmoins, c'est l'addition des trois réponses apportées aux limites décrites à la Figure 3-1 qui permet de raffiner l'approche pedigree et ainsi de répondre à l'objectif global de ce projet de recherche.

Les sections qui suivent ne proposent qu'un survol des cadres méthodologiques à développer pour répondre aux questions et aux objectifs de recherche. Les chapitres suivants offrent une perspective plus détaillée sur la méthodologie mise en œuvre dans le cadre de ce travail ainsi que sur les hypothèses posées, elle y est présentée sous la forme d'articles scientifiques.

3.3.1 Légitimer l'usage de la distribution log-normale comme distribution par défaut

Afin de tester l'impact du choix de la distribution log-normale comme distribution par défaut pour modéliser une donnée et son incertitude, il est proposé d'analyser des cas de figures spécifiques à l'ACV (plutôt que de proposer une réponse théorique à cette question).

Les analyses d'incertitudes menées en ACV vont notamment permettre de discriminer un scénario par rapport à un autre et ainsi de choisir celui avec le moins d'impacts potentiels sur l'environnement avec une quantification du niveau de confiance à accorder à cette décision. Ainsi, le cadre d'analyse développé ici doit permettre à la fois de tester l'impact sur les résultats ACV obtenus par analyse d'un système de produit mais aussi de le tester sur des comparaisons entre différents systèmes. Pour ce faire, les points présents à la Figure 3-2 sont repris.

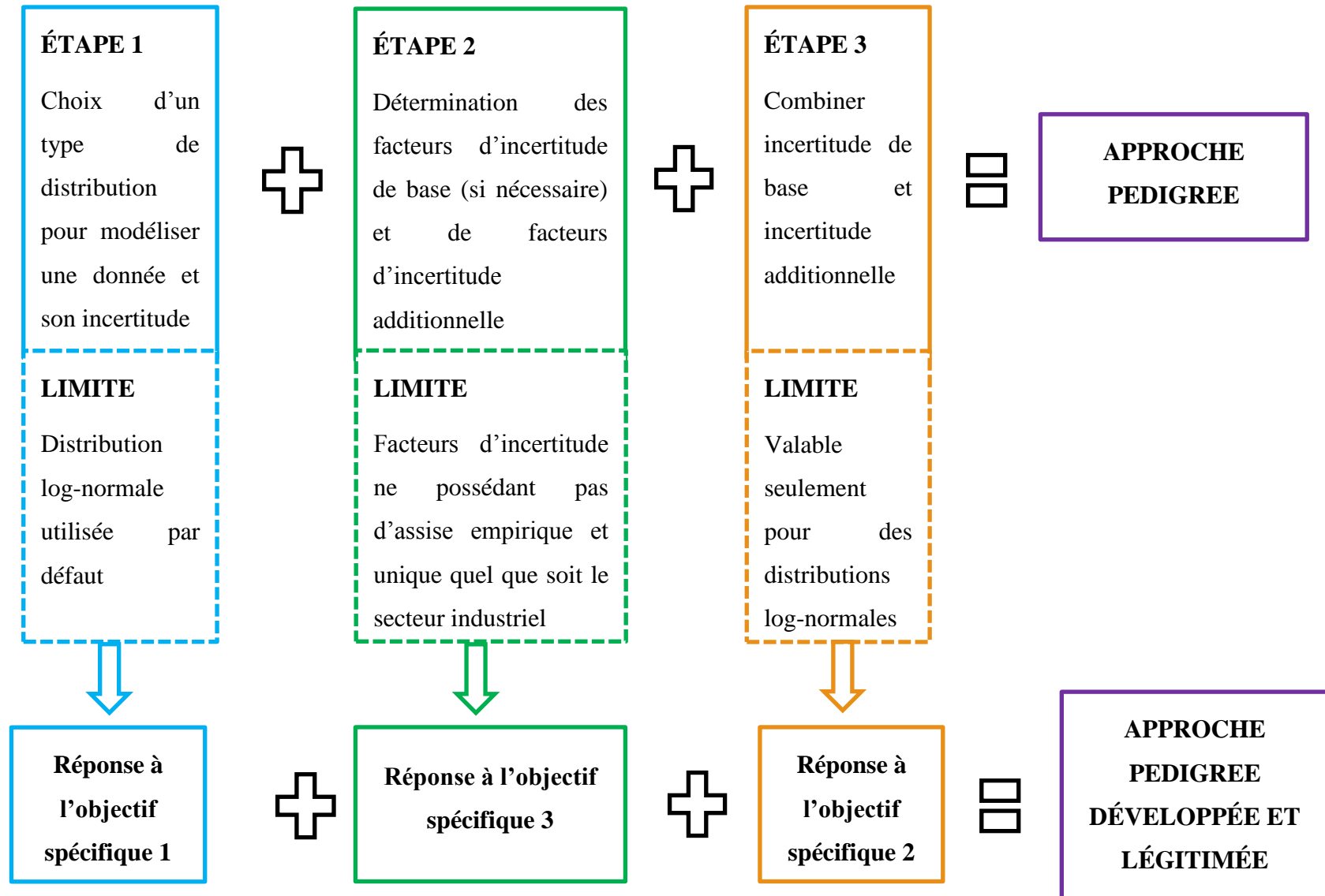


Figure 3-1 : Étapes de l'approche pedigree, limites associées et traitement de ces dernières pour répondre à l'objectif de recherche

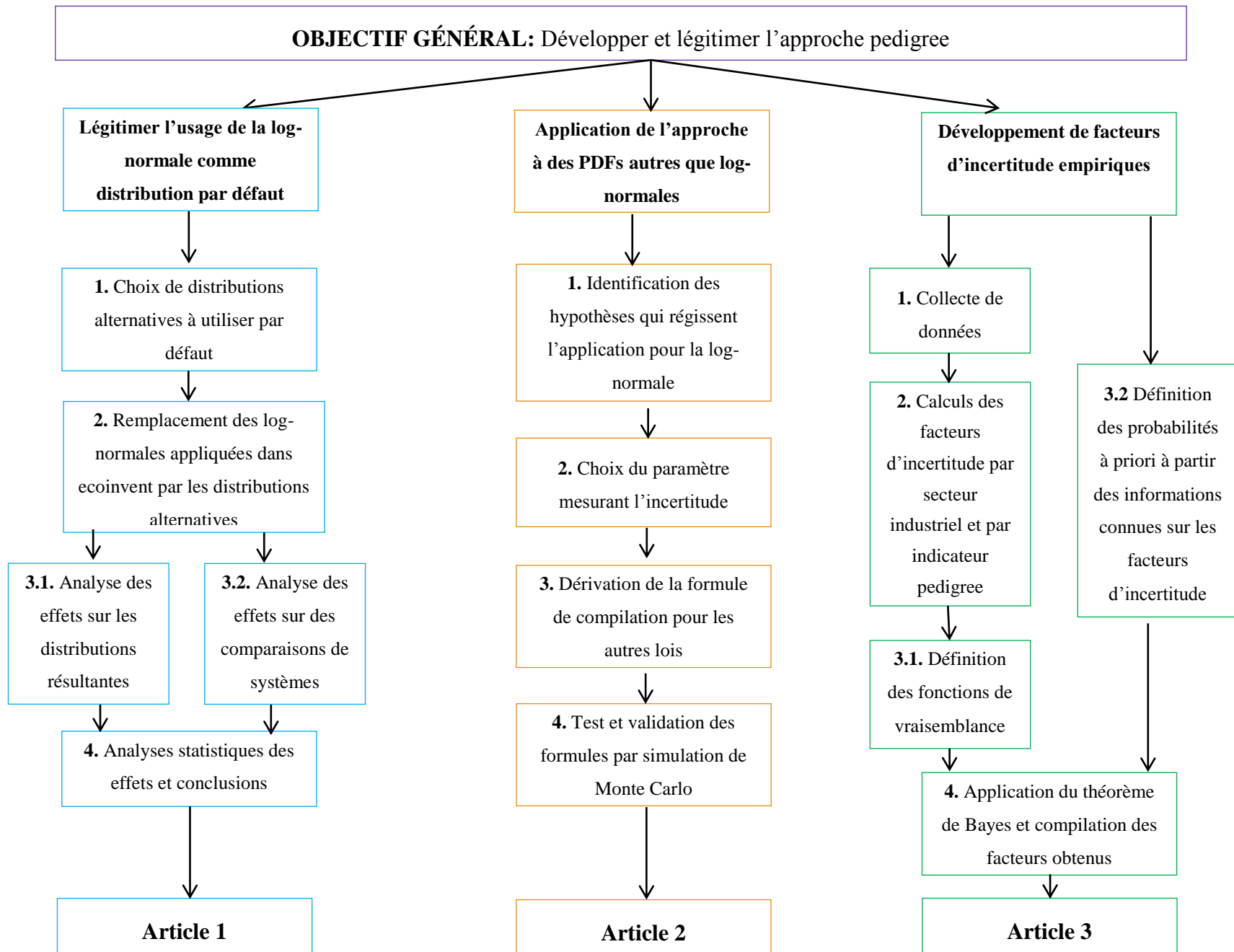


Figure 3-2 : Schéma méthodologique général

1. Choix de distributions alternatives à utiliser par défaut

Les distributions à tester pour le remplacement des distributions log-normales sont choisies parmi les distributions disponibles dans la base de données ecoinvent (cf. leur définition au Tableau 2-3). Elles sont considérées afin qu'un changement d'intervalle de support de la distribution (intervalle non borné vs. intervalle borné) ainsi qu'un changement dans l'asymétrie de la distribution (distribution asymétrique vs. distribution symétrique) soient testés, et ce, de manière indépendante. Une distribution ayant les mêmes caractéristiques que la distribution log-normale (à intervalle de définition non borné et à asymétrie positive) est également testée pour le changement. Ainsi les distributions choisies sont les suivantes :

- La distribution gamma pour ses caractéristiques similaires à la log-normale;
- La distribution triangulaire comme distribution définie sur un intervalle borné;
- La distribution normale pour sa similarité avec la log-normale et le fait qu'elle soit symétrique (si une variable aléatoire X suit une distribution log-normale, alors la variable aléatoire $Y=\ln(X)$ suit une distribution normale).

2. Remplacement des distributions log-normales présentes dans la base de données ecoinvent par les distributions alternatives précédemment choisies

Afin de pouvoir remplacer l'ensemble des lois log-normales modélisant une donnée et son incertitude par une des trois distributions précédemment choisies, il s'agit de déterminer les paramètres de chacune de ces distributions en se basant sur les paramètres de définition de la log-normale. Ainsi, pour les lois normale et gamma, les égalités de la médiane et de l'écart-type sont conservées. Pour ce qui est de la loi triangulaire, le cas diffère puisque les degrés de liberté entre les paramètres ne sont pas les mêmes. Des hypothèses sont donc posées : l'égalité des valeurs déterministes (la médiane pour la log-normale et le mode pour la triangulaire) est conservée et la détermination du minimum et du maximum de la loi triangulaire est basée sur la définition de l'intervalle à 95% de la loi log-normale.

3. Analyse des effets du remplacement

i. *Sur les paramètres des distributions résultantes*

Une fois les distributions remplacées, un calcul d'ÉICV est réalisé sur les systèmes de produits modélisés dans la base de données ecoinvent v2.2. Ce calcul d'ÉICV est réalisé sur l'impact potentiel de la catégorie « Changements climatiques » pour les quatre distributions et sur l'ensemble des systèmes de produits.

Le calcul d'incertitudes permettant d'analyser l'effet des choix se fait en utilisant les techniques d'analyse de Monte-Carlo, le résultat de cette analyse de Monte-Carlo est donc une distribution de probabilités modélisant l'impact potentiel sur les changements climatiques d'un système de produit avec son incertitude. Les effets du changement de distribution sur la distribution résultante sont déterminés par rapport aux distributions résultantes obtenus lorsque la distribution log-normale est utilisée par défaut et est calculée en utilisant le calcul de ratios. Ces ratios sont déterminés pour différents paramètres de la distribution résultante : la médiane, le coefficient de variation et le coefficient d'asymétrie.

ii. *Sur les comparaisons entre systèmes de produits*

Pour comparer les effets du changement lors de la comparaison de deux systèmes de produits, ces derniers sont choisis dans la base de données ecoinvent v2.2. de sorte à ce qu'ils modélisent des types de produits et d'activités différents. 160 systèmes de produits sont conservés et comparés deux à deux pour l'ensemble des quatre distributions. Afin d'effectuer cette comparaison, il s'agit dans un premier temps de s'affranchir des effets d'échelle qui surviennent lorsque des activités différentes sont comparées. Dans un second temps, la probabilité (notée x_{ij}) pour la différence entre les deux systèmes d'être positive est calculée puis compilée pour l'ensemble des comparaisons obtenues, et ce, pour les quatre distributions.

4. Analyses statistiques des effets et conclusions

Pour conclure sur les effets du changement de distributions sur les systèmes de produits, une analyse statistique des ratios compilés est menée. En ce qui a trait aux comparaisons entre les

systèmes de produits, les valeurs x_{ij} obtenues pour chaque distribution sont également comparées par le biais du calcul puis de l'analyse des ratios obtenus.

3.3.2 Application de l'approche pedigree aux autres types de distribution présentes dans ecoinvent v3

Par application de l'approche pedigree, on entend ici la compilation de l'incertitude de base et incertitude additionnelle pour obtenir le facteur d'incertitude totale. Cette compilation n'est ni justifiée ni expliquée dans les documents expliquant la prise en compte de l'incertitude dans ecoinvent (Frischknecht et al., 2005; Weidema & Wesnæs, 1996).

1. Identification des hypothèses qui régissent l'application de l'approche pedigree pour une distribution log-normale utilisée par défaut

Afin de pouvoir appliquer l'approche aux autres lois disponibles dans la base de données ecoinvent v3, il s'agit dans un premier temps de comprendre son fonctionnement pour la loi log-normale. Ce fonctionnement se base sur les faits suivants :

- L'ajout de l'incertitude additionnelle ne modifie ni la valeur déterministe (i.e. la valeur utilisée pour définir le flux), ni le type de distribution utilisée pour représenter l'incertitude;
- Lorsque aucune incertitude additionnelle n'est ajoutée (i.e. quand tous les scores pedigree sont égaux à 1), l'incertitude totale est égale à l'incertitude de base;
- L'incertitude additionnelle est exprimée de manière adimensionnelle et représente la dispersion relative liée à chaque score pedigree et qui est due à l'utilisation d'une donnée imparfaite pour le contexte de l'étude;
- Les facteurs de conversion des scores pedigree aux facteurs d'incertitude additionnelle (cf. Tableau 2-9) sont utilisés pour la détermination de l'incertitude additionnelle.

2. Choix du paramètre permettant l'expression du facteur d'incertitude

Plusieurs paramètres exprimant le facteur d'incertitude ont été proposés et utilisés en lien avec l'approche pedigree. Dans l'article définissant l'approche pour la première fois, Weidema et Wesnaes (1996) proposent le CV pour exprimer le facteur d'incertitude. Dans la deuxième version d'ecoinvent, le carré de l'écart-type géographique (GSD^2) est utilisé (Frischknecht et al., 2005). Dans la troisième version de la base de données le carré de l'écart-type logarithmique est quant à lui utilisé (σ_{LN}^2). L'écart-type logarithmique est également l'écart-type de la loi normale associée à la loi log-normale : si la variable aléatoire X suit une loi log-normale de paramètres μ_{LN} et σ_{LN} alors la variable aléatoire $Y=\ln(X)$ suit une loi normale de paramètres μ_{LN} et σ_{LN} . Pour la version 3 d'ecoinvent, les GSD^2 ont donc été convertis en termes de σ_{LN}^2 en se basant sur le lien existant entre les deux paramètres pour la loi log-normale ($GSD = \exp(\sigma_{LN})$) (Weidema et al., 2013).

Il s'agit ici de choisir le paramètre permettant d'exprimer les facteurs d'incertitude et d'appliquer l'approche pedigree pour l'ensemble des distributions, ce paramètre se doit donc d'être :

- Adimensionnel : permettant ainsi de l'appliquer quels que soient l'unité et l'ordre de grandeur du flux considéré;
- Facilement interprétable pour l'ensemble des distributions.

3. Dérivation de la formule de compilation des incertitudes de base et additionnelle pour les autres types de distribution

La dérivation se base à la fois sur l'ensemble des hypothèses qui ont été définies au point 1 et la traduction de ces dernières à la fois pour le paramètre défini au point 2 et l'Équation 2-1 combinant incertitudes de base et adimensionnelle.

4. Tests et validation des formules obtenues par simulations de Monte-Carlo

Afin de tester l'adéquation des formules de compilation obtenues pour l'obtention du facteur d'incertitude totale par application des formules, ce dernier est comparé à la valeur obtenue par simulation numérique. L'analyse de Monte Carlo est également utilisée ici et les différences entre

la compilation de l'incertitude de base et incertitude additionnelle pour la méthode développée et par simulation numérique sont étudiées en analysant les spécificités des deux distributions résultantes modélisant un flux avec son incertitude totale.

3.3.3 Développement de facteurs empiriques pouvant être mis à jour

1. Collecte de données

Afin de déterminer de nouveaux facteurs d'incertitude empiriques pour les données utilisées en ICV, une collecte de données est nécessaire. Ces données doivent obéir à un certain nombre de critères afin de pouvoir être utilisées dans ce contexte précis :

- L'ensemble des typologies de données utilisées en ICV doit être représenté : les données collectées doivent représenter et les flux élémentaires et les flux économiques;
- Les données doivent être exprimées par unité fonctionnelle ou par unité d'activité;
- Les sources de données utilisées doivent couvrir un large spectre de secteurs industriels, d'années d'obtentions et d'aires géographiques, ainsi que de manières dont les données ont été acquises afin d'avoir suffisamment de données pour chaque couple « Indicateur pedigree, score pedigree ».

Pour obtenir ce type de données, des rapports d'ACV, des articles scientifiques traitant d'ACV et d'ICV, des bases de données spécifiques et des bases de données de facteurs d'émissions⁵ sont utilisés.

2. Calcul des facteurs d'incertitude par secteur industriel et par score pedigree

Une fois les données compilées dans une base de données spécifique à ce projet et préparées afin de pouvoir les analyser sur une base commune, celles-ci sont classées par couple « indicateur pedigree, score pedigree » en suivant les définitions de la matrice (cf. Tableau 2-7) et par secteur industriel. Cette classification en secteurs industriels se base sur la classification industrielle

⁵ Les facteurs d'émissions permettent de déterminer la valeur d'une émission à l'environnement en connaissant le « taux » de l'activité étudiée (p.ex. la masse d'un produit fabriqué). Les facteurs d'émissions sont souvent utilisés pour déterminer les quantités de pollutions rejetées par une activité ou une entreprise.

nord-américaine (NAICS - <https://www.census.gov/eos/www/naics/>). Un classement sur le seul couple « indicateur, score » permet, quant à lui, de développer des facteurs d'incertitude génériques.

Pour chacun des sous-groupes ainsi formés, le facteur d'incertitude est calculé en se basant sur les travaux préalablement réalisés par Ciroth et collègues (2013).

La construction des sous-groupes pour la détermination des facteurs d'incertitude de base est un peu différente. En effet, l'incertitude de base peut être considérée comme l'incertitude résultante lorsque tous les scores pedigree sont égaux à 1 (i.e. quand la donnée utilisée est parfaite pour le contexte de l'étude). Dans ce cas, les sous-groupes sont donc créés à partir de données provenant d'une même technologie, d'une même année, d'un même pays et pour lesquelles le score pour le critère « justesse » vaut 1.

Le cas particulier de l'indicateur « complétude » : Si l'on se rapporte aux définitions pour ce critère présentes au Tableau 2-7, il peut être vu comme le critère mesurant la représentativité des données par rapport aux sites considérés ou en d'autres termes à la représentativité d'un échantillon par rapport à la population que cet échantillon se doit de représenter. Afin de déterminer les facteurs d'incertitude propres à ce critère, un lien avec les statistiques propres aux techniques d'échantillonnage est réalisé et utilisé pour déterminer la variabilité due à cet échantillonnage.

3. Définition des éléments permettant l'application du théorème de Bayes

Les facteurs pedigree tels qu'appliqués dans les versions 2 et 3 de la base de donnéesecoinvent proviennent de jugement d'experts. Ce jugement est une connaissance en soi qu'il est possible de combiner à une nouvelle information empirique obtenue. Cette combinaison peut se faire en appliquant le théorème de Bayes tel que présenté à la section 2.3.2.2. Dans ce projet, la connaissance sur les facteurs d'incertitude émanant des jugements d'experts est conservée car l'information concernant les facteurs d'incertitude obtenue empiriquement reste partielle. En effet, l'ensemble des sources de données collectées manuellement au point 1 ne permet pas de représenter l'ensemble des échanges au sein de la technosphère (i.e. les échanges entre les processus unitaires) ni les échanges entre la technosphère et l'écosphère (i.e. les échanges entre les processus unitaires et l'environnement).

i. Détermination des distributions de probabilités à priori

Avant de déterminer la distribution à priori des facteurs d'incertitude, il s'agit de recenser ces derniers et de les exprimer selon une mesure commune de l'incertitude. Des facteurs d'incertitude issus de jugements d'experts sont présents dans l'article de Weidema et Wesnaes (1996) et dans l'article présentant la base de données ecoinvent v2.2 (Frischknecht et al., 2005). À ceux-ci s'ajoutent les facteurs obtenus par Ciroth et collègues préalablement à ce projet et publié dans un article (2013) et dans un rapport remis à l'*ecoinvent centre* (2012). Ces données permettent de définir les fonctions à priori pour chacun des facteurs d'incertitude. Puisque, suivant les facteurs d'incertitude, 1 à 26 données sont disponibles pour définir ces distributions, des hypothèses cohérentes pour l'ensemble des facteurs sont posées pour développer les distributions à priori.

ii. Détermination des fonctions de vraisemblance

La fonction de vraisemblance est une fonction f qui lie les données aux paramètres d'intérêt. Le paramètre d'intérêt est ici le facteur d'incertitude (UF) et les données sont les facteurs d'incertitude empiriques obtenus au point 2, ils peuvent donc être décrits par $f(UF)$ où UF est la moyenne de la fonction de vraisemblance. Cette fonction f est donc déterminée à partir des facteurs empiriques obtenus au point 2.

4. Application du théorème de Bayes, compilation des facteurs obtenus et possibilité de mise à jour de ces facteurs

Suivant la nature des fonctions de vraisemblance et des distributions à priori, le théorème de Bayes peut soit être appliqué de manière théorique soit par le biais de simulations numériques (Ben Letham, 2012; Qian, et al., 2003). Pour l'application du théorème dans ce cas, fonctions de vraisemblance et distributions à priori sont définies pour être conjuguées, le théorème peut donc s'appliquer de manière théorique.

L'application du théorème de Bayes permet d'obtenir une fonction à posteriori des facteurs d'incertitude dont la moyenne peut être considérée comme les facteurs mis à jour et combinant jugements d'experts et information empirique. Cette distribution à posteriori peut être utilisée comme distribution à priori si de nouveaux facteurs d'incertitudes empiriques sont disponibles et

le théorème de Bayes peut être appliqué une deuxième fois pour obtenir une nouvelle mise à jour des facteurs d'incertitude, et ainsi de suite.

CHAPITRE 4 ARTICLE 1: EFFECTS OF DISTRIBUTION CHOICE ON THE MODELLING OF LCI UNCERTAINTY – AN ASSESSMENT ON THE ECOINVENT V2.2 DATABASE

4.1 Présentation de l'article

Le manuscrit qui suit présente le travail effectué lors d'un stage à Zürich (Suisse) à l'ETH (*Eidgenössische Technische Hochschule* – l'Institut fédéral de technologie) au sein du groupe ESD (*Ecological system design* – design des systèmes écologiques). Il a été rédigé en collaboration, en plus des directeurs de recherche, avec Christopher Mutel qui a également supervisé le stage.

Il a été **soumis** le 12 novembre 2015 au sein du *Journal of Industrial Ecology*. Les informations supplémentaires soumises en même temps que le manuscrit se trouvent à l'Annexe B.

Ce manuscrit recense les conclusions quant à la justification de l'utilisation de la distribution log-normale comme distribution par défaut pour modéliser un flux avec son incertitude totale, et ce, pour l'approche pedigree.

4.2 Manuscrit

4.2.1 Introduction

4.2.1.1 The life cycle assessment calculation model

Life cycle assessment (LCA) is defined as a holistic, systematic and objective tool that aims to quantify all the potential environmental impacts generated by a product or a service from resource extraction to end-of-life (Jolliet, 2005; Junnila & Horvath, 2003; Lo et al., 2005). This information on a product's environmental performance can be used by decision-makers in the fields of policy making, production, consumption and marketing. LCA relies heavily on modelling, from the defining of a product's system to the definition of the environmental cause-effect chain.

In order to determine all these potential impacts, a matrix model can be used (for complete description of this model, see Heijungs and Suh (2002)). This model links:

- A technological matrix **A** that contains the amount of product exchange needed or produced by each unit process.
- An intervention matrix **B** that contains the amount of elementary flow emitted or consumed by each unit process;
- An impact matrix **Q** that contains the impact characterisation factors of all elementary flows for each impact category;
- A final demand vector **f** representing the reference flow of a LCA.

These elements are linked to obtain the vector of impacts **h**.

$$\mathbf{h} = \mathbf{QBA}^{-1}\mathbf{f} \quad \text{Equation 4-1}$$

4.2.1.2 Uncertainties in LCA and life cycle inventories

Uncertainty analysis, which is a part of the LCA methodology, conveys information to the decision maker on the robustness of the results (Huijbregts, 1998; Maurice et al., 2000).

There are a number of uncertainty types in all four phases of a LCA. These are commonly classified as model, scenario and parameter uncertainty (Lloyd & Ries, 2007; Reap et al., 2008). There is no universal approach to consider uncertainties at large in LCA, and uncertainty results cannot be compared from one study to another (Weber, 2012). Approaches include, among others, sensitivity, scenario and uncertainty analyses. A review by Lloyd and Ries (2007) shows that inventory parameter uncertainty is considered the most often, and we will consider only inventory parameter uncertainty in this paper.

Different methods have been proposed to manage parameter uncertainties in life cycle inventory (LCI). Three main categories of methodologies can be cited: qualitative, semi-quantitative and purely quantitative. The most common way to quantify and model uncertainty on parameters is the use of statistical tools and in particular classical statistical theory to derive probability density functions (PDFs) (Finnveden et al., 2009). The effect of the uncertainties on the final results can

be quantified using uncertainty propagation methods such as Monte Carlo simulation or analytical methods based on Taylor series expansion (Ciroth et al., 2004; Hong et al., 2010; Imbeault-Tétreault et al., 2013). Many widely used LCA software tools (e.g. SimaPro, GaBi, CMLCA, OpenLCA, Brightway2) offer the possibility to run uncertainty analysis through Monte Carlo simulation.

4.2.1.3 The use of probability density functions to model parameter uncertainties in LCI

The derivation of PDFs to model an LCI parameter and its uncertainty can be based on statistical methods if the available sample is large enough. Most of the time, due to data and resource constraints, expert judgment or generic factors are used to choose a PDF type (Clavreul et al., 2013; Lloyd & Ries, 2007).

In order to overcome the bias that can arise when practitioners choose different PDFs in a same study (Geisler, et al., 2005), some LCA developers proposed the use of default distributions to model a LCI parameter and its uncertainty. For example, the beta distribution can model have a large variety of shapes and ranges (Canter et al., 2002; Kennedy et al., 1996; Wang & Shen, 2013). A particular form of the beta distribution, the beta PERT, was proposed by Koffler and Kalish (2012). The beta PERT is less flexible than the beta distribution but can be more easily defined (see Tableau 4-1 for a census of the distributions found in LCI and their main characteristics). To our knowledge, the beta and beta PERT distributions have only been used in case studies, e.g. the assessment of beverage delivery systems (Canter et al., 2002; Kennedy et al., 1996) or the estimation of the embodied energy of major building materials (Wang & Shen, 2013). In the ecoinvent database, the lognormal distribution is chosen as the default (Frischknecht et al., 2005). Using the lognormal distribution and a semi quantitative approach to manage uncertainties allows the ecoinvent database to be the only LCI database to quantify and model uncertainty on nearly all flows present in the database and hence for all background data used in a LCA.

Besides these three distributions, other – more marginal – distributions were used in LCI to model uncertainty on a parameter; Tableau 4-1 makes a census of these distributions with their characteristics and the type of flows they modeled.

Tableau 4-1 Characteristics of probability distribution and their use in LCI

Distribution	Definition parameters	Main characteristics	Use in LCI
Lognormal	Logarithmic mean and standard deviation (the mean and the standard deviation of the underlying normal distribution)	Nonnegative Unbounded – interval $[0, +\infty[$ Positively skewed Captures a large range	As default for all type of LCI parameters in the ecoinvent database (Frischknecht et al., 2005) To model pollutant releases (Daniels & Higgins, 2002; Hamid et al., 2013; Sonnemann et al., 2003) To model energy consumption and emissions based on best fitting tests (Silva, 2011).
Normal	Mean and standard deviation	Unbounded – interval $]-\infty, +\infty [$ Symmetric	Alternative to lognormal default in ecoinvent v3 (Weidema et al., 2013) To model intermediate exchanges (Sonnemann et al., 2003) To model emissions and activity data (Röös et al., 2010)
Gamma	Shape and scale parameters – location parameter (0 by default)	Unbounded – interval $[0, +\infty[$ Positively skewed – less heavy tail than the lognormal distribution	Alternative to lognormal default in ecoinvent v3 database (Weidema et al., 2013) To model PM distributions (Hamid et al., 2013) To model vehicle emissions, based on real samples (Lloyd & Ries, 2007)
Weibull	Shape and scale parameters – location parameter (0 by default)	Unbounded - interval $[0, +\infty[$ Positively skewed – less heavy tail than the gamma distribution	To model energy consumption and emissions based on best fitting tests (Silva, 2011).
Beta	Location parameters (minimum and maximum) – two shape parameters	Bounded Great variety of shapes; depending on the shape parameters the distribution can be symmetric, positively or negatively skewed.	Used by default in global framework due to its large variety of shape (Canter et al., 2002; Kennedy et al., 1996; Wang & Shen, 2013; Wang et al., 2012)
Uniform	Location parameters (minimum and maximum)	Bounded Symmetric Outside the bounds, the probability directly falls to 0	Alternative to lognormal default in ecoinvent v3 database (Weidema et al., 2013) Use when parameters are not studied enough or when there is a debate around them (Lloyd & Ries, 2007; Maurice et al., 2000; Mullins et al., 2010)
Triangular	Location parameters (minimum and maximum) and mode	Bounded Some flexibility in the shapes, depending on the value of the mode Intuitive distribution that overemphasises the tails	Alternative to lognormal default in ecoinvent v3 database (Weidema et al., 2013) Use when parameters are not studied enough or when there is a debate around them and when the mode is known (Lloyd & Ries, 2007; Maurice et al., 2000; Mullins et al., 2010)
Beta PERT	Location parameters (minimum and maximum)	Bounded	Alternative to lognormal default in ecoinvent v3 database (Weidema et al., 2013)

	and mode	<p>Some flexibility in the shapes, depending on the value of the mode</p> <p>Less sensitive to the extreme values than the triangular distribution</p>	<p>et al., 2013)</p> <p>Used in an industrial project to model parameter uncertainty on scenarios (Koffler & Kalish, 2012)</p> <p>Used in a case study when assessed data have low quality (Maurice et al., 2000)</p>
--	----------	--	---

The different PDFs can be categorized through their range and their shape:

- *Bounded vs. unbounded distributions*: the definition domain of bounded distribution is finite and described by a minimum and a maximum value. Unbounded distributions are defined on an infinite or semi-infinite domain, although extreme values are associated with correspondingly low probabilities of occurrence.
- *Symmetric vs. skewed distributions*: The skewness of a distribution is defined by its third standardized moment γ_1 and it is a measure of asymmetry (or, in other terms, of central tendency). If the skewness parameter is equal to zero, then the distribution is symmetric and the most likely value, median and mean are all equal. For other cases, the sign of the skewness parameter indicates the shape of the distribution. For positive skewness (the lognormal distribution for example), the majority of the data lies around the left of the mean and extreme values to the right.

4.2.1.4 The use of the lognormal distribution as default distribution

The lognormal distribution is the default distribution applied in the ecoinvent database. For this specific application, the lognormal distribution has several useful properties. First, it is guaranteed to always be positive, which is convenient when modelling physical quantities (Frischknecht et al., 2005). Next, it is linked to the normal distribution by the following property: the random variable X is lognormally distributed if the variable $\log(X)$ is normally distributed. The normal distribution is known to model physical quantities resulting from the addition of independent processes. Similarly, the lognormal distribution models variables resulting from the multiplication of independent processes (Bebbington & Zitikis, 2008; Limpert et al., 2001). Because they are driven by multiplicative processes, many variables in natural science can be modelled with a lognormal distribution. Daniels and Higgins (2002) and Limpert et al. (2001) provide some examples; among them, pollutant concentrations (Morgan, et al., 1992).

Furthermore, the lognormal distribution is quite good to model large uncertainties due to its skewness (Limpert et al., 2001). Finally, its definition parameters (the logarithmic mean and standard deviation) can be defined even if only a few data points or dataset parameters are available.

In the ecoinvent v2.2 database, over 99.9% of uncertain LCI parameters are lognormally distributed. How does the choice of this default distribution affect the uncertainty of the resulting indicator results? The characteristics of the resulting impact assessment score distribution cannot be predicted when performing an uncertainty assessment. The goal of this work is to measure the effect of changing default distribution of LCA parameters inputs by assessing the resulting distribution of outputs and by measuring the robustness of comparisons between two product systems.

4.2.2 Materials and methods

In order to assess the effects of a change in the default distribution on the impact assessment results, the lognormal distributions will be switched to alternative distributions, based on the methodology presented below.

In the database, the uncertainty information is stored in the uncertainty field in the ecoSpold data files. This field contains, in case of the lognormal distribution, the deterministic value (the median); the geometric standard deviation as a measure of intrinsic uncertainty; the five pedigree indicators representing the additional uncertainty (for more information on the pedigree matrix used in ecoinvent, see Weidema and Wesnæs (1996) and Muller et al. (2014)); and the total geometric standard deviation that takes into account the intrinsic and additional uncertainty. The measures of interest here are the deterministic value and the total geometric standard deviation; they allow the definition of the PDF that models a LCI parameter with its total uncertainty. This uncertainty information will be used in order to perform the switch from the lognormal to another default distribution.

4.2.2.1 Selection of distributions and conversion from lognormal to alternative default PDFs

Alternative distributions are chosen from the parameter distributions available in ecoinvent version 3.

- A distribution with **the same characteristics as the lognormal** (unbounded and positively skewed). The gamma distribution is positively skewed and encompasses the same interval $([0, +\infty[)$ as the lognormal distribution;
- Positively skewed and **bounded default distribution**. The triangular distribution is selected, as its parameters are easily understandable and definable;
- Unbounded and **symmetric default distribution**. The normal distribution is selected due to its similarities with the lognormal distribution. For a sample with a coefficient of variation (ratio between the standard deviation and the mean) smaller than 0.2 a normal and a lognormal distribution can be indifferently used to model the sample (Morgan et al., 1992).

For the normal and the gamma distribution, the switch is straightforward. The equations linking the median and standard deviation of the lognormal distribution and the mean and standard deviation of the normal distribution are known. The link between the lognormal median and standard deviation and the shape and the scale parameters of the gamma distribution can also be uniquely determined by solving a quadratic system (see Tableau 4-2).

The approach used for the triangular distribution is a bit different. This distribution can take only a limited number of shapes (Karlis, 2008). In order to derive the parameters, it was so decided to set the mode c of the triangular equal to the median value of the lognormal distribution, as these are the “most representative” values for these distributions. The minimum and maximum values, a and b , are based on the 2.5th and the 97.5th percentiles of the lognormal distribution.

All the assumptions and parameter definitions can be found in Tableau 4-2.

Tableau 4-2 Determination of the definition parameters for the distribution used for the switch

Distribution	Probability density function	Definition parameters	Deterministic value (as defined in ecoinvent)
Lognormal	$f(x, \mu, \sigma) = \frac{\exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})}{x\sigma\sqrt{2\pi}}$	μ and σ : the logarithmic mean and standard deviation	Median = $\exp(\mu)$
Normal	$f(x, \mu_N, \sigma_N) = \frac{\exp(-\frac{(x - \mu_N)^2}{2\sigma_N^2})}{\sigma_N\sqrt{2\pi}}$	μ_N and σ_N : the mean and standard deviation $\mu_N = \exp(\mu)$ $\sigma_N = \exp(\mu)\sqrt{\exp(\sigma^2) - 1}$	Median = mode = mean = μ
Gamma	$f(x, k, \theta) = \frac{x^{k-1} \exp(-x/\theta)}{\Gamma(k)\theta^k}$ for $x \geq 0$	Where k is the shape parameter and θ the scale parameter $\theta = \frac{-\exp(\mu) + \sqrt{\exp(2\mu) + 4 * (\exp(\mu + 0.5 * \sigma^2) * \sqrt{\exp(\sigma^2) - 1})^2}}{2}$ $k = \left(\frac{\exp(\mu + 0.5 * \sigma^2) * \sqrt{\exp(\sigma^2) - 1}}{\theta} \right)^2$	Mode = $(k-1)\theta$
Triangular	$\begin{cases} f(x, a, b, c) = \frac{2(x-a)}{(b-a)(c-a)} \\ \text{for } a < x < c \\ f(x, a, b, c) = \frac{2(b-x)}{(b-a)(c-a)} \\ \text{for } c < x < b \\ f(x, a, b, c) = 0 \text{ otherwise} \end{cases}$	Where a is the minimum, b the maximum and c the mode of the distribution $c = \exp(\mu)$ $a = \frac{\exp(\mu)}{\exp^2(\sigma)}$ $b = \exp(\mu) \exp^2(\sigma)$	Mode = c

4.2.2.2 LCA calculations and results comparison

1. Assessing the effects on parameters

Once the distributions are switched, a life cycle impact assessment (LCIA) is performed on product systems from the ecoinvent 2.2 database using the IPCC 2001 (100 year timeframe) characterization factors. The LCIA is performed for 3981 product systems modelled with the four

versions of the database; we exclude product systems where the climate change category is zero or where the product systems are aggregated. For each of the 3981 product systems and four databases distributions, a 5000 iteration Monte Carlo impact assessment uncertainty analysis is performed. The results are themselves distributions of impact scores, which are referred to as the **resulting distribution** in the following sections.

The statistical parameters of the resulting distributions, representing their central tendency, their dispersion and their skewness, will be affected by the default distribution used in the database. In order to assess the effects of the distribution switch, the statistical parameters of the resulting distribution are compared to the default situation (i.e. using lognormal distributions as default). The comparison is simply a ratio, where the statistical parameter of the resulting distribution when alternative distributions are used as default is divided by the statistical parameter obtained when the lognormal distribution is used. These ratios permit to determine the relative size of the differences between resulting statistical parameters as well as the direction of the effect. The comparison is done on the median, standard deviation and skewness of the resulting distributions for each product system.

2. Assessing the effects on product systems comparison

This section describes the procedure used to evaluate the effect of the switch on product system comparison. It was decided not to compare all 3981 product systems. First, comparing each possible pair of product systems would generate an overwhelming amount of data (there are almost 8 million possible comparisons). Also, given that some product systems are very similar (e.g. different truck transport systems), part of the information would be redundant. The selection of product systems is based on the processes' classification used SimaPro software. The top-level categories are material, energy, transport, processing, use, and waste treatment, each of which is divided in several subcategories (total of 160 subcategories). One product system was randomly chosen from each – see the supporting information for the complete list.

The 160 products systems are then compared to each other for every default distribution based on their climate change indicators. This results in 12,800 pair-wise comparisons per distribution type. For each comparison, the compared product systems are generically referred to as P_i and P_j .

The comparison is described here for the lognormal case, and is repeated with data using the alternative default distributions.

A 5000 step Monte Carlo analysis is carried out where calculate climate change indicator results are simultaneously calculated for all product systems P . The resulting indicator results are stored in 160 vectors $\mathbf{p}_{i,LN}$, each containing 5000 results, where the subscript LN refers to the lognormal distribution used as default. The k^{th} element in any vector $\mathbf{p}_{i,LN}$ is based on the same sampling as the k^{th} element in another vector $\mathbf{p}_{j,LN}$.

At this moment, the climate change indicator results for two product systems P_i and P_j , contained in $\mathbf{p}_{i,LN}$ and $\mathbf{p}_{j,LN}$, cannot sensibly be compared since their outputs can be wildly different (for example, P_i can represent a transport system with an output expressed in tkm and P_j can represent a metal production system with an output expressed in kg). In order to obtain, artificially, product systems that can be compared, the vectors $\mathbf{p}_{i,LN}$ and $\mathbf{p}_{j,LN}$ are normalized by their respective medians. The resulting normalized vectors of climate change indicator results, respectively referred to as $\mathbf{p}'_{i,LN}$ and $\mathbf{p}'_{j,LN}$, now both have the same median. However, their respective probability distribution functions are a function of the Monte Carlo simulation and are not identical.

The normalized distribution $\mathbf{p}'_{j,LN}$ is then multiplied by a factor $\alpha=0.9$. This artificially creates a situation where the median of the compared systems' indicator results are different by a factor of 10% which, according to Humbert et al. (2009), is the rule-of-thumb minimum difference in deterministic climate change indicator results required to consider two product systems significantly different. In other words, if uncertainty analysis were not carried out, and assuming the median to be a good estimation of the deterministic result, the two product systems would be considered to be significantly different for the climate change impact category. Note that other shifting factors α could have been chosen, and results for a range of α values are presented in the Supporting information.

The indicator of interest is the probability that the product system $\alpha\mathbf{p}'_{j,LN}$ indeed has a climate change indicator lower than $\mathbf{p}'_{i,LN}$. This is calculated as the relative frequency of $(p_{i,LN,k} - \alpha p_{j,LN,k}) > 0$ over all iterations k .

The same methodology is then applied using the alternative default distributions. This allows the comparison of the probability that the product system $\alpha\mathbf{p}'_j$ has a climate change indicator lower

than P'_i for different default uncertainty distributions. Tableau 4-3 presents this methodology in a mathematically formalised way.

Tableau 4-3 Methodological scheme to assess the effects of the used distribution of product systems comparison

<i>Notation</i>	<p>$\mathbf{p}_{i,LN}$: the vector containing the 5000 outputs resulting from the Monte Carlo simulation for the i_{th} product system and the lognormal distribution (LN).</p> <p>The subscripts N, G and T are respectively used for the normal, gamma and triangular distribution.</p>
Step 1	<p>Normalisation of the product system:</p> <p>$\mathbf{p}'_{i,LN} = \mathbf{p}_{i,LN} / \text{median}(\mathbf{p}_{i,LN})$ (respectively for $\mathbf{p}_{i,N}$, $\mathbf{p}_{i,G}$, and $\mathbf{p}_{i,T}$)</p>
Step 2	<p>For every couple (i,j)</p> <p>$x_{ij,LN} = p(\mathbf{p}_{i,LN} - \alpha \mathbf{p}_{j,LN} > 0)$ with $\alpha=0.9$ in the baseline calculation (respectively for $x_{ij,N}$, $x_{ij,G}$, and $x_{ij,T}$)</p>
Step 3	<p>Determination of the ratio on the probability x_{ij}</p> <p>$r_{ij,LN-N} = x_{ij,LN} / x_{ij,N}$ (respectively for $r_{ij,LN-G}$ and $r_{ij,LN-T}$)</p>
Step 4	Assessing the different ratios obtained

4.2.3 Results and discussion

4.2.3.1 Effects on individual product systems

In this section, the effects of switching default distributions is assessed on the resulting distribution of the climate change indicator result for each of the product systems. The assessment is performed on the distribution characteristics: the central tendency through the

median, the range through the standard deviation and the shape through the coefficient of skewness. Some of the presented results are available in the supporting information.

1. Switch effect on the median

The choice of default distribution affects the central tendency of the resulting distribution. This effect is assessed by comparing, for each of the 3981 product systems and each of the four default distributions, the ratio between the resulting median and the deterministic result (see Figure 4-1).

For the gamma and the triangular distribution, the resulting median is equal to or higher than the deterministic result (i.e. considering uncertainty increases the central tendency of the indicator result). The same conclusion can be drawn for the lognormal distribution except for 77 product systems. For the normal distribution, most ratio values are slightly smaller than 1, i.e. considering uncertainty slightly decreases the central tendency.

In order to more specifically assess the consequence of changing the default distribution from the lognormal to the other three distributions, Figure 4-2 presents the same results, but using the lognormal median as the denominator for the ratio instead of the deterministic result. The effect on the central tendency is confirmed: switching to the normal distribution decreases the median and switching to the gamma or the triangular distribution increases the median, and this effect is most important for the triangular distribution.

The way the definition parameters of each distribution are calculated, i.e. based on the parameters of the lognormal distribution, can explain these differences. The equality of the deterministic value was conserved rather than the equality of e.g. the median (see Tableau 4-2). And, while the median is the deterministic value for the lognormal and the normal distributions, the mode is the deterministic value for the gamma and the triangular distributions. Note that in the ratios calculations, the mean – rather than the median - could also have been used as a measure of the central tendency.

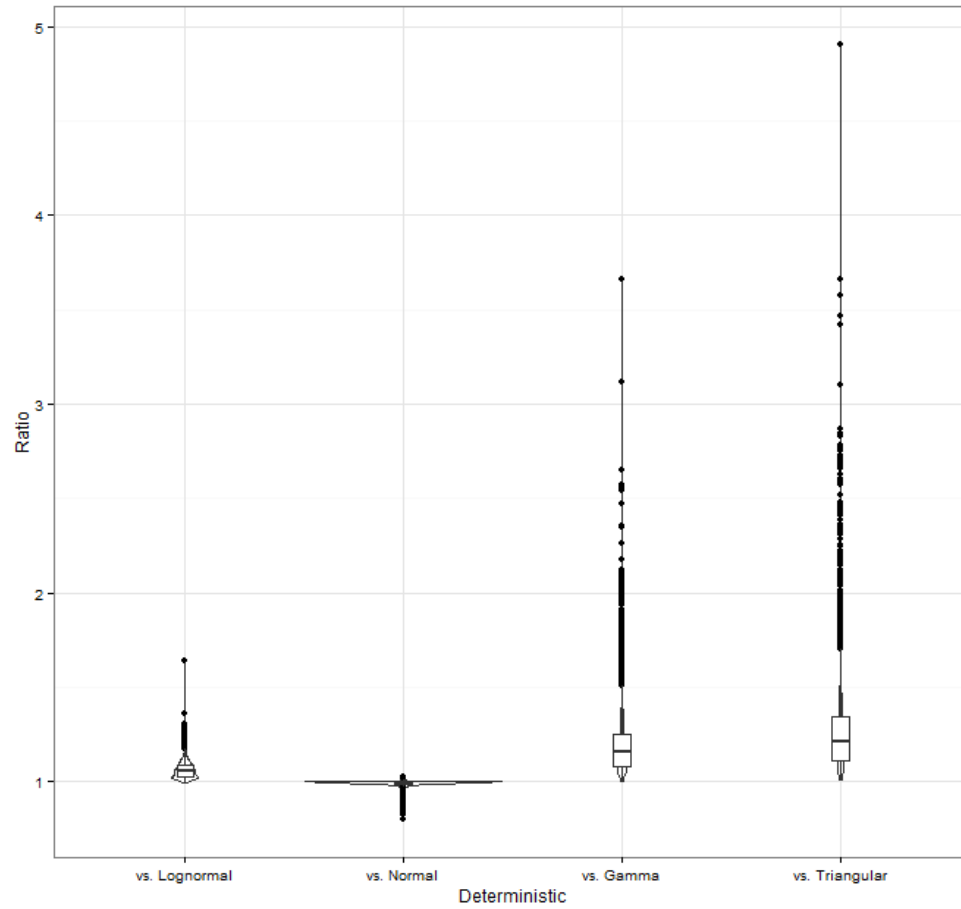


Figure 4-1 Effect of the choice of the default distribution on the central tendency, expressed as the ratio between the median of the resulting distribution and the deterministic value and for each assessed default distribution used to model uncertainty of inputs

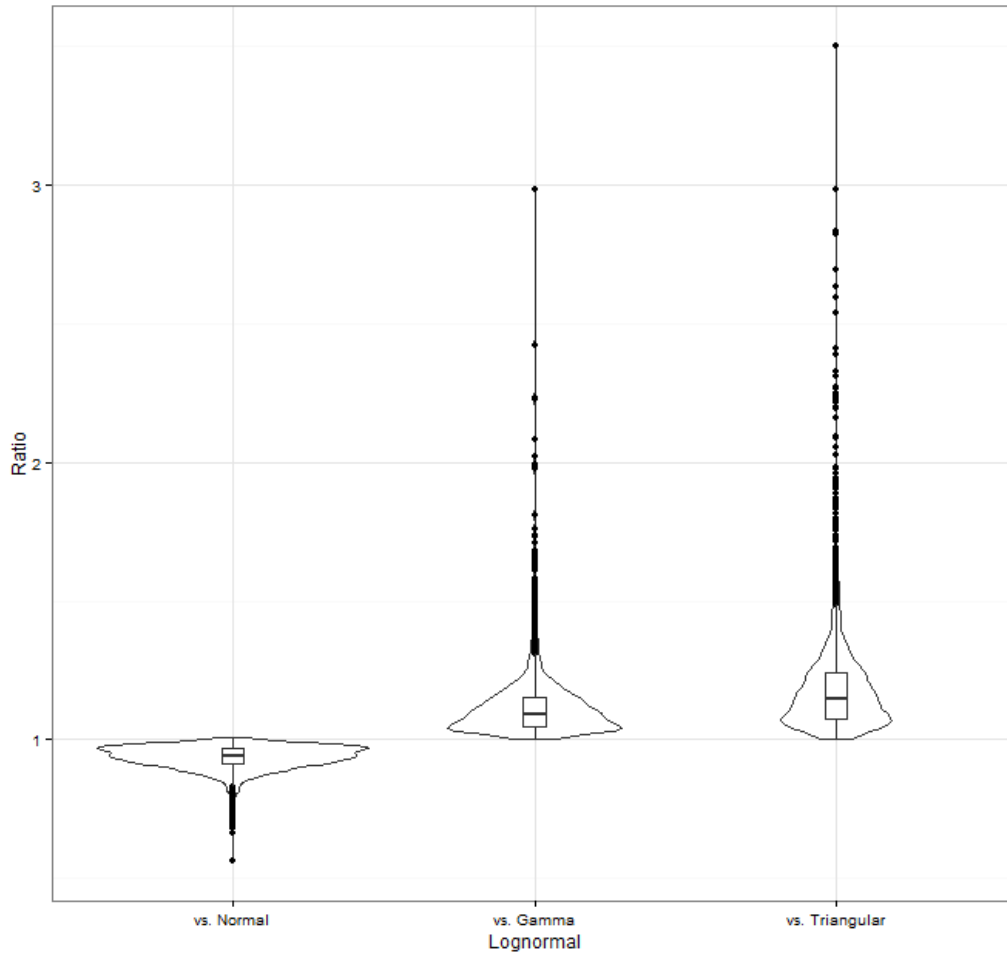


Figure 4-2 Effect of a switch in choice of default distribution on the central tendency, expressed as the ratio between the median of the resulting distribution when an alternative default distribution is used and when the lognormal distribution is used as default

2. Switch effect on the variability

The effect of the choice of distribution on the variability is assessed by taking the ratio of the standard deviation of the resulting distributions calculated using an alternative distribution (normal, gamma or triangular) and using the default distribution (lognormal). These ratios, presented in Figure 4-3, are generally smaller than 1 for the normal and the triangular distribution, showing that the standard deviations of the resulting distribution are greater when the lognormal distribution is used. This is especially stark for the normal distribution, where 99% of the standard deviations are higher than when the lognormal is used as default. In the case of

the gamma distribution, all but 250 ratios are smaller than 1, showing that the standard deviations of the resulting distributions are generally higher when the gamma distribution is used. For 5 product systems, the standard deviation obtained with the lognormal distribution is several orders of magnitude higher than those calculated with other distributions. These specific products systems have the particularity to have relatively high uncertainty defined on the input parameters (i.e. the defined uncertainty factors of the unit processes supplying the reference flow are higher than 2).

These observations can be linked to the distributions characteristics and to the calculations of the definition parameters for each tested distribution. The fact that the triangular distribution is bounded and defined based on the lognormal 95th confidence interval explains the smaller variability. The equality of the variance for the gamma and lognormal distribution when defining the gamma definition parameters explains the small differences in the resulting variability. Finally, even though the normal distribution parameters are defined based on the variance equality with the lognormal distribution, the symmetry of the normal distribution produces resulting distributions that are less skewed (see Figure 4-4), which can explain the differences in the resulting distribution variability.

3. Switch effect on the skewness

The skewness of each resulting distribution is presented in Figure 4-4. The differences are very significant for all distribution comparisons. The skewness measures the asymmetry of a distribution. The majority of the resulting distributions (for each default distribution) are positively skewed: most of the values lie near the deterministic value and the distribution is right-tailed.

The differences are more pronounced when the use of the normal and triangular distributions are compared with the use of the lognormal as a default distribution. These observations can be explained by the characteristics of the distributions. For the normal distribution, the skewness parameter equals 0, i.e. the distribution is symmetric. For the triangular distribution, the observations can be explained by the definition of the skewness based on the location parameters only and on the fact that this distribution can only take a limited number of shapes (Karlis, 2008).

Finally, as with the lognormal, the gamma distribution is unbounded and positively skewed, explaining the smaller differences on the resulting skewness.

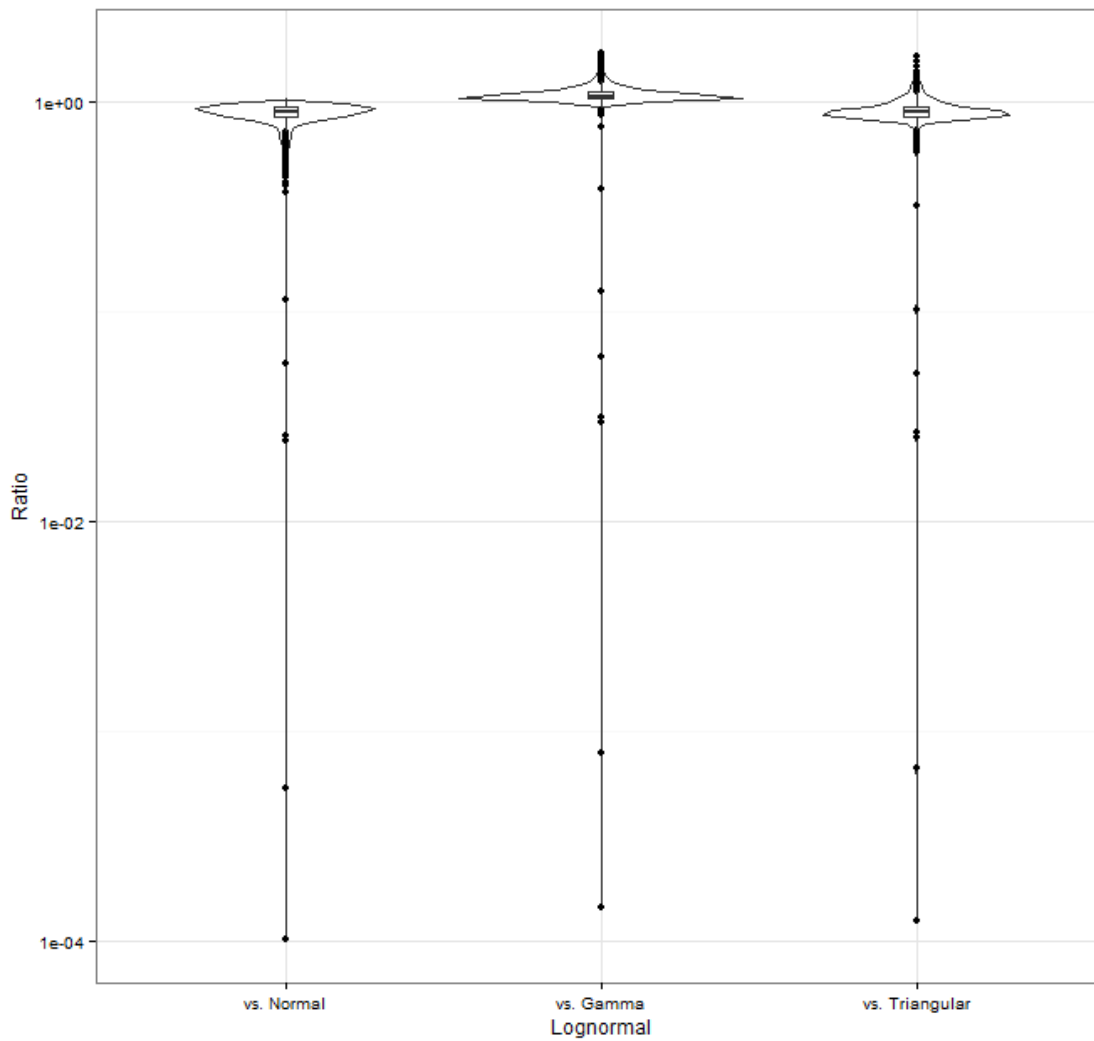


Figure 4-3 Effect of a switch in choice of default distribution on dispersion, expressed as the ratio between the standard deviation of the resulting distribution when an alternative distribution is used by default and when the lognormal distribution is used as default (logarithmic scale)

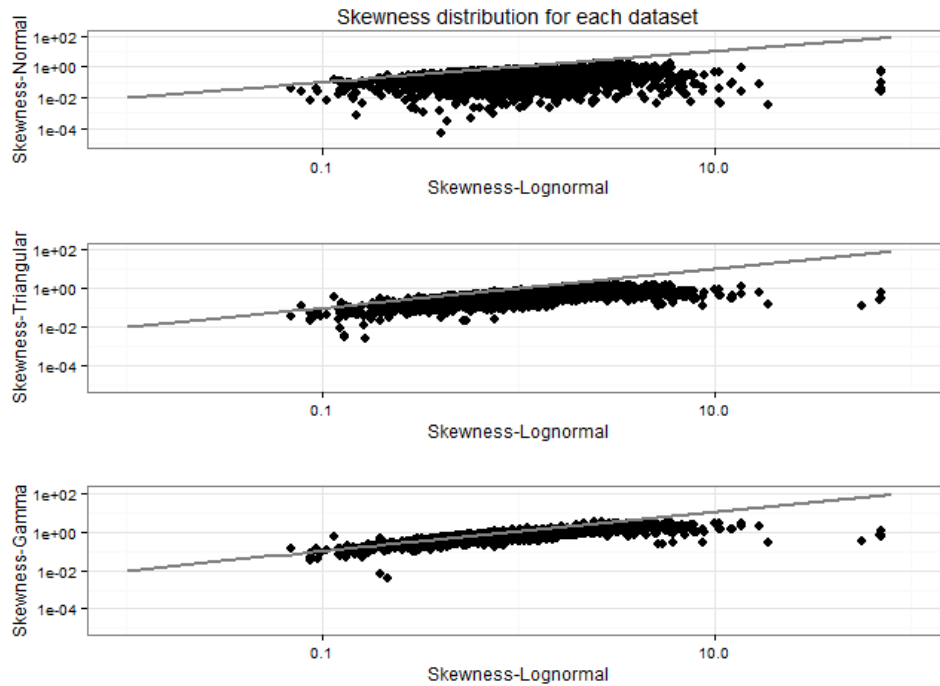


Figure 4-4 Skewness distribution γ_1 for each product system – on logarithmic scale

4. Sumamry

Figure 4-6 shows the obtained trends on the resulting distributions after the switch for a specific product system. These trends are summarised in Tableau 4-4, which shows the conclusions when comparing the resulting distribution obtained with the lognormal used by default and each of the three switches. We observe a generally positive translation of the resulting median when asymmetric distributions are used as default, which can be explained by the fact that the mode is the defined deterministic value for both the triangular and the gamma distribution. The variability on the final result is generally greater when the lognormal distribution is used as default. Finally, for all comparison the resulting distribution is more skewed when the lognormal distribution is used by default. The use of the lognormal distribution leads to an emphasis on the high values of the resulting distribution. The different calculated results are available in the supporting material.

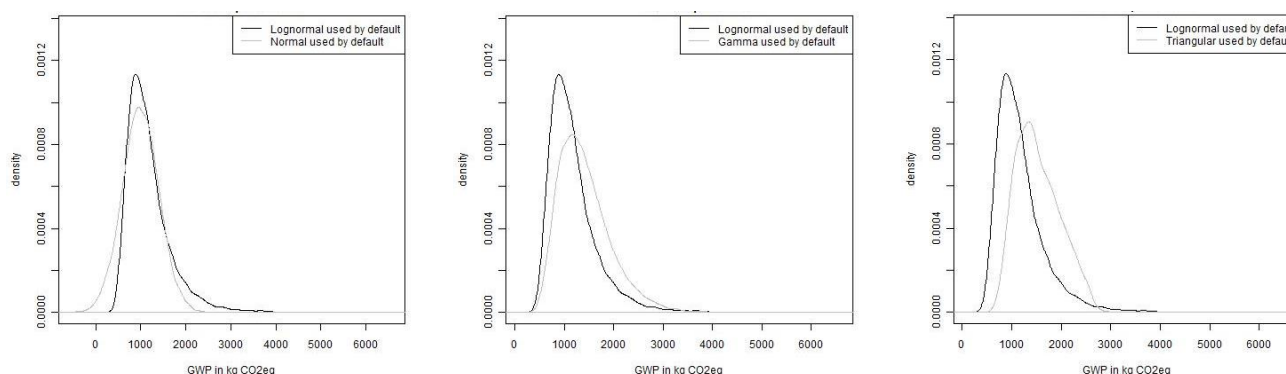


Figure 4-5 Resulting distribution for the climate change impact for the three switch and for the same product systems

Tableau 4-4 Obtained trends in the comparison on the resulting distributions between the lognormal distribution used by default and each of the three switches. Ratios are calculated taking the lognormal distribution as a numerator

	Normal	Triangular	Gamma
Distributions characteristics	Unbounded and symmetric	Bounded and asymmetric	Unbounded and asymmetric
Resulting distribution median	Low differences on the median Mean of ratio: 0.94	Low translation of the median Mean of ratio: 1.2	Low translation of the median Mean of ratio: 1.1
	General conclusion: $\text{Normal} \leq \text{Lognormal} \leq \text{Gamma} \leq \text{Triangular}$		
Resulting distribution variability expressed through the standard deviation	Some differences Mean of ratio: 0.88	Some differences Mean of ratio: 0.91	Low differences Mean of ratio: 1.01
	General conclusion: $\text{Normal} \leq \text{Triangular} \leq \text{Lognormal} \leq \text{Gamma}$		
Resulting distribution skewness	Almost symmetric distribution.	Less skewed distribution than when	Less skewed distribution than when

		the gamma and the lognormal are used by default.	the lognormal is used by default.
	General conclusion: Normal \leq Triangular \leq Gamma \leq Lognormal		

4.2.3.2 Effects on product systems comparison

Figure 4-6 shows an example of the distribution of the difference between two product systems (P_i and P_j). These differences are calculated when the lognormal distribution is used by default and for the three alternative default distributions. In the default situation, the difference has a 81.4% probability to be positive, i.e. $0.9p_{j,LN,k}$ has a 81.4% probability of having a score that is lower than that of $p_{i,LN,k}$. This probability is nearly the same when the normal and the gamma distribution are used by default (the probabilities are respectively 80.4% and 81.0%); when the triangular distribution is used by default, this probability increases to 86.0%

The same calculation is made for the 12,720 product systems comparison, leading to a mean of ratios between the lognormal and the normal distribution ($r_{ij,LN-N}$) of 1.01, 1.00 between the lognormal and the gamma distribution ($r_{ij,LN-G}$) and 0.96 between the lognormal and the triangular distribution ($r_{ij,LN-T}$). The distributions of these ratios are illustrated in Figure 4-7. As these ratios are nearly 1.00, a switch of the distribution doesn't sensibly affect a comparison between two products systems. The same conclusions can be drawn when the factor α used to translate one of the resulting distributions varied from 0 to 2, (see supporting information).

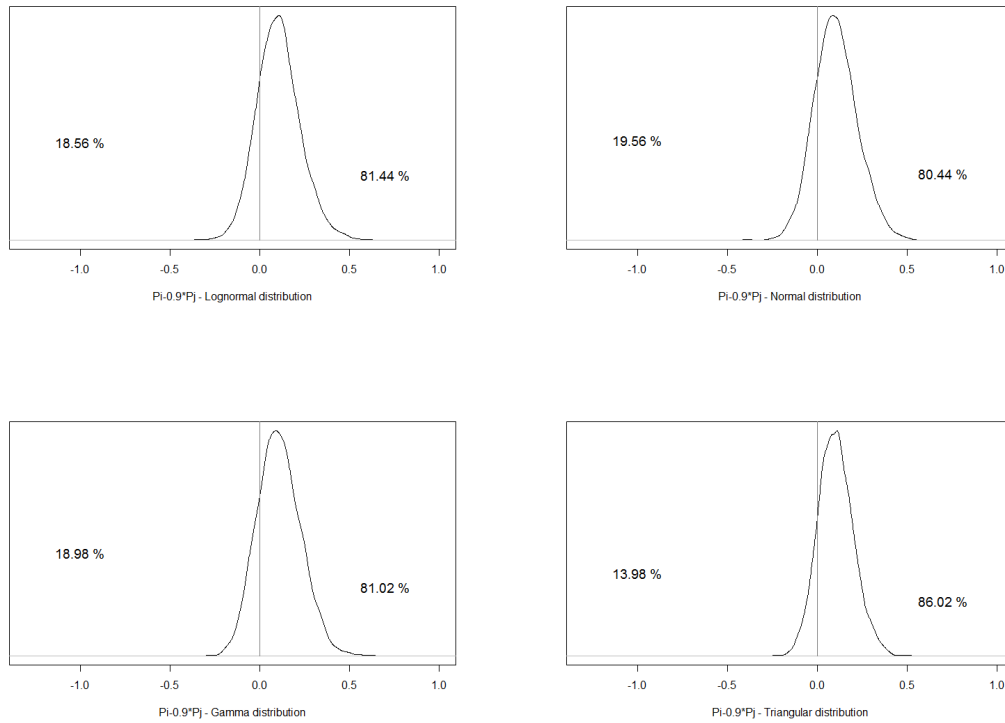


Figure 4-6 Monte Carlo results for the difference between two product systems (P'_i and $\alpha P'_j$). Percentages indicate the probability for the result to be negative or positive.

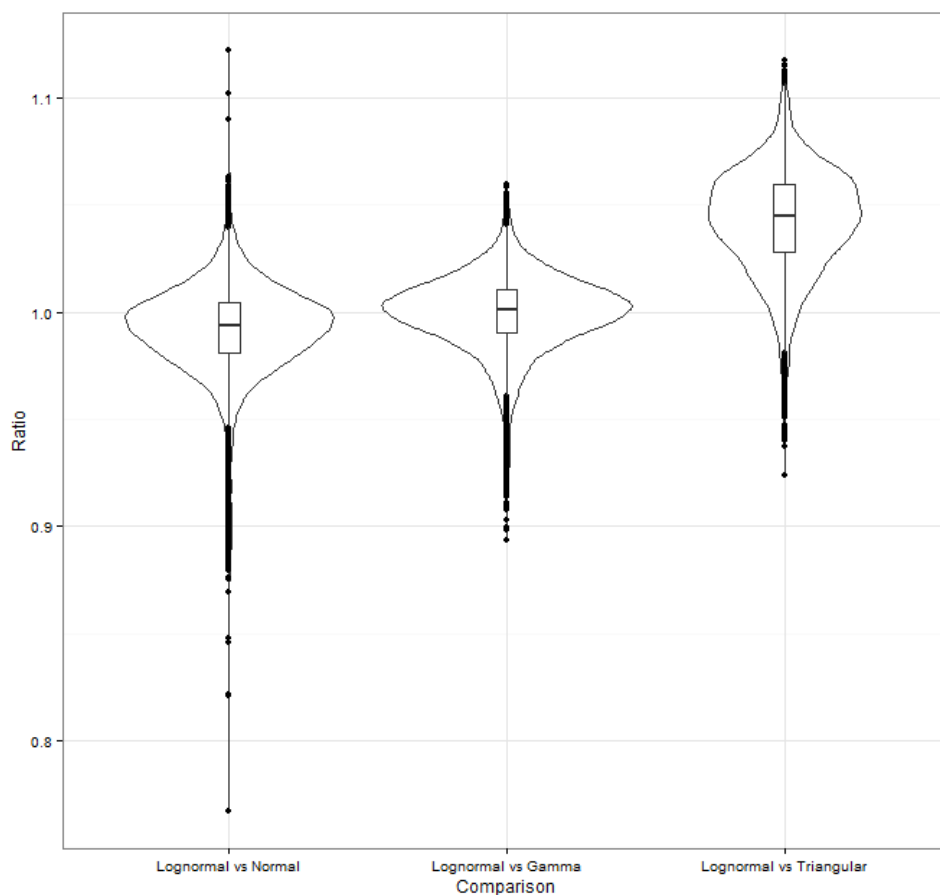


Figure 4-7 Ratio on the product system comparison, lognormal distribution as a denominator

4.2.4 Limits and recommendations

4.2.4.1 Limitations of the “switch” model

In order to simulate an ecoinvent v2.2 database where another distribution had been used as default, a conversion from reported lognormal distributions to other distributions was needed. This substitution is based on some choices and assumptions.

First assumption concerns how the central tendency is defined. In this study, the deterministic value as expressed in the ecoinvent database is chosen as the central tendency on which the normal, gamma and triangular distributions are constructed (see Tableau 4-2). This assumption is made because, in ecoinvent, the deterministic value is used as the representative of the modelled LCI parameter with its uncertainty and is set equal to the median of the lognormal distribution.

The results of this study would certainly change if another measure of central tendency were chosen (for example if the mean were chosen rather than the ecoinvent deterministic value). For positively skewed distributions (like the lognormal, the gamma and, in this specific case, the triangular distributions), the following relation holds: $\text{Mean} > \text{Median} > \text{Mode}$. This implies that if the mean were chosen rather than the deterministic value, the median of the resulting distributions would be higher. This would also affect the resulting variability and skewness, since the definition parameters of the distributions also define the standard deviation and the skewness of the distribution (see Tableau 4-1). For the normal distribution (and other symmetric distributions), the equality $\text{Mean} = \text{Median} = \text{Mode}$, and so the results would not change based on how the central tendency is defined.

The second assumption concerns how dispersion is defined. When possible, the equality of the variance was conserved. It was not the case for the triangular distribution, where the dispersion was determined through the lognormal 95th percentile confidence interval. As seen, the lognormal distribution is positively skewed with a long right-tail, so the choice of the 95th percentile confidence interval seemed a good compromise here. The 99th percentile confidence interval was also tested; leading to triangular distribution with a variance much larger than the one that was modelled with a lognormal distribution (the difference between both variances varies exponentially and can reach two orders of magnitude when the original logarithmic standard deviation varies from 0.5 to 1.5).

This study is also limited to the use of three alternative default distributions with different characteristics. Other distributions have been used in the field of LCA (see Tableau 4-1) and would have been interesting to look into (the beta PERT for example).

Despite these assumptions and choices, this analysis of the effect of a switch can be seen as a sensitivity assessment around the use of default distribution to model data uncertainty. The conclusions of this study shows that the characteristics of the resulting distribution depends on the distribution used to model uncertainty inputs. However, the switch has practically no effect on product systems comparison.

4.2.4.2 Uncertainty modelling choices affect resulting distributions

The goal here is not to recommend a type of distribution to use by default to model uncertainty of background data (for foreground data, if information on uncertainty is available, it should be used to model the LCI parameter with its uncertainty). Rather, it is to show the differences and some trends when comparing the use of default distributions to model uncertainty on data. The effects on the resulting distributions are shown in relation to the use of the lognormal distribution. The most obvious conclusion, which was predictable, is that the skewness of results is greater when the lognormal is used as the default distribution. This fact should be kept in mind when interpreting the results of an uncertainty assessment that uses the lognormal distribution by default.

This fact, and the differences pointed out in the results section, show that there is a need to be aware that some characteristics in the resulting distributions are linked to the default distribution used to model uncertainty; especially on the median and on the skewness.

In the same way, due to these differences, it is recommended to use the same default distribution to model uncertainty on data when comparing two products systems. It has been shown that switching distributions has no effect on the conclusion of a product systems comparison; however using different by default distributions in the two systems might lead to differences in the resulting distribution due to the distributions characteristics and not on the uncertainty itself. These conclusions can be applied in other fields than LCA when the same kind of models as the one described in Equation 4-1) are used (for example in ecotoxicology).

Finally, the work here underlines the possibility of more flexibility in LCA modelling: combining LCA data and some language programming led to more complex and advanced LCA calculations that allowed enhanced interpretation.

CHAPITRE 5 ARTICLE 2: THE APPLICATION OF THE PEDIGREE APPROACH TO THE DISTRIBUTIONS FORESEEN IN ECOINVENT V3

5.1 Présentation de l'article

L'article présenté ici a été élaboré dans le cadre du « projet pedigree » (cf. section 3.1) qui visait à légitimer l'approche pedigree telle qu'appliquée dans la base de données ecoinvent avant la sortie de la troisième version de la base de données. Dans ce contexte et en plus des directeurs de recherche, cet article a été rédigé en collaboration avec :

- Andreas Ciroth, en tant que principal acteur de la première phase du « projet pedigree »;
- Christopher Mutel, en tant que réviseur des lignes directrices pour la version 3 de la base de données d'ecoinvent, et ce, notamment pour le chapitre touchant à la modélisation de l'incertitude;
- Bo P. Weidema, en tant qu'« inventeur » de l'approche pedigree telle qu'appliquée en ICV et représentant de l'*ecoinvent centre*.

Il a été **publié** dans une édition spéciale de l'*International Journal of Life Cycle Assessment* portant sur la base de données ecoinvent v3 après avoir été accepté le **16 mai 2014**. Le matériel supplémentaire publié avec l'article est, quant à lui, disponible dans l'Annexe C du présent document.

Les résultats présentés dans cet article permettent d'appliquer l'approche pedigree pour l'ensemble des distributions permettant la modélisation d'une donnée avec son incertitude dans ecoinvent v3. Plus spécifiquement, les résultats obtenus permettent de combiner incertitude de base et incertitude additionnelle afin de modéliser l'incertitude de la donnée.

5.2 Manuscript

5.2.1 Introduction

LCA is a quantitative model of the industrial ecosystem, and LCA results always have uncertainty, i.e. are not perfectly known. Sources and typologies of uncertainty in LCA have been well documented (Huijbregts et al. 2003; Lloyd and Ries 2007; Björklund 2002; Sonnemann et al. 2003). Uncertainty can be divided into model, scenario and parameter uncertainty. The focus of this paper is on the parameter uncertainty due to lack of knowledge associated to the true value of a quantity) associated with the values of exchanges (elementary and intermediate flows) in the context of Life cycle inventory (LCI) databases. Although we recognize the importance of uncertainty in life cycle impact assessment, this paper discusses only the inventory phase.

These parameters, the quantitative data needed to compile a LCI (e.g. exchange values), are uncertain for a number of reasons; for example, the real value could not be known (Björklund 2002), or the measured data do not accurately describe the “true” median or mean value (Ciroth et al. 2004). In real-world technological and environmental processes, there is also considerable variability. The concept of variability is often confused with uncertainty. Variability refers to the inherent variations of a parameter value in space, time or across individuals (Huijbregts, 1998) while uncertainty (specifically, parameter uncertainty) includes also the lack of confidence that a parameter value is truly representative of the true value. It is important to consider uncertainties in the interpretation phase of LCI and LCA studies in order to determine the confidence in results. Usually, a quantitative parameter is known with a certain level of uncertainty and a certain level of quality. Taking into account this uncertainty and quality of the inputs in a LCA study, especially on parameters, will reinforce the confidence in the results and help the decision-making process based on the results and their interpretation in a quantitative and qualitative way. Several approaches were also developed in order to take these uncertainties at a parameter level, see Lloyd and Ries (2007) for a review, including purely qualitative (Rousseaux et al. 2001), semi-quantitative (Kennedy et al. 1996; Maurice et al. 2000; Weidema and Wesnæs 1996) and quantitative (Tan et al. 2002; Ciroth et al. 2004; Coulon et al. 1997) approaches. Different approaches were also proposed to consider how the parameter-level uncertainty translates into uncertainty of actual LCA results (Hong et al. 2010; Ciroth et al. 2004), although this topic will for the most part be outside the scope of this paper.

The LCI phase of an LCA aims to collect and compile all the heterogeneous data needed to quantify the exchanges within the studied product system and between the product system and the environment. Since product systems model complex supply chains, LCA practitioners and researchers normally revert to using generic LCI databases that contain exchange values for different processes. Since these databases are an important data source for LCA, how they describe uncertainty in their data is the key to subsequently evaluating the overall uncertainty of the LCA.

The ecoinvent database is one of a very few LCI databases that systematically includes explicit uncertainty data. A semi-quantitative approach based on the use of a pedigree matrix is used. This approach, first proposed for the LCA by Weidema and Wesnaes (1996), is inspired from the NUSAP system developed in the early 90's (Funtowicz and Ravetz 1990), and has been in use in the ecoinvent database since 2005 (Frischknecht et al. 2005). It considers two kinds of parameter uncertainty:

- 1) intrinsic variability and stochastic error of the parameters, due to e.g. measurement uncertainties, activity specific variability, temporal variability, etc. This uncertainty is captured in a so-called *basic uncertainty* factor;
- 2) uncertainty due to the use of imperfect data, e.g. data resulting from estimates, lacking verification, or extrapolated from temporally, spatially and/or technologically different conditions, collectively grouped under the term *additional uncertainty*.

In other terms, the approach addresses two of the three categories of parameter uncertainty described in Huijbregts et al. (2001), namely data inaccuracy and lack of representative data for the context of use. In the approach, a datum and its uncertainty (the basic and additional uncertainty) is modeled using a probability density functions (PDF). In the earlier version of the ecoinvent database, the lognormal distribution was used by default (Frischknecht et al. 2005). The lognormal distribution is represented by two definition parameters: the geometric mean (μ_g) and the geometric standard deviation (GSD) see Tableau 5-1. The geometric mean (also the median) is the deterministic value and the GSD captures the information on the uncertainty. Uncertainty factors were also expressed in terms of GSD^2 (basic uncertainty factors) and “contributors to the GSD^2 ” (additional uncertainty factors) in ecoinvent v2. Having statistical information about the data, basic uncertainty factors were directly calculated as GSD^2 . When no

statistical information was available, basic uncertainty factors could be defined by default. For this, ecoinvent provided a table with basic uncertainty factors differentiated by exchange type (total of 27 types, ranging from specific classes of air pollutants to different types of intermediate exchanges) and by class of process (combustion, process or agricultural). The datum was then evaluated according to a “pedigree matrix” by which it was scored (1 to 5, where 5 is worse) on the following independent data quality characteristics: reliability (sampling methods and verification procedures); completeness (statistical representativeness of the datum and time periods for data collection); temporal, geographic and further technological correlation (for data used outside its proper context) and sample size.

The semi-quantitative scores, based on the pedigree matrix, were then converted into uncertainty figures using additional uncertainty factors, expressed as a contribution to the square of the geometric standard deviation. For example, a reliability score of “5”, representing a value based on a non-qualified estimate, would be associated with a contributor to the GSD^2 of 1.50. These additional uncertainty factors were subsequently compiled with the basic uncertainty factors using a compilation formula, derived from the properties of the lognormal distribution, resulting in a measure of the total uncertainty. The total uncertainty was expressed as the square of the geometric standard deviation (GSD^2). If in ecoinvent, the “pedigree matrix approach” refers to the determination of the additional uncertainty factors; the term “pedigree approach” will be used in this paper to name the whole semi quantitative structure: from the determination of the additional uncertainty to the compilation of both types of uncertainty.

In the context of a database, this approach has many advantages. Uncertainty information can be made available for all types of exchanges, even if statistical information is not available. This ensures that the cumulative uncertainty of LCIs, calculated via, e.g., Monte Carlo or analytical propagation approaches, takes into account the uncertainty of every exchange in the model. The data quality indicators derived from the use of the pedigree matrix can also be used directly as a data quality management tool.

Tableau 5-1 Definition of the distributions foreseen in ecoinvent v3 and how to express the CV according to the definition parameters

Name of the distribution	Parameters	Deterministic value	PDF	How to link the coefficient of variation (CV) with the parameters?
Lognormal	μ_g : geometric mean σ_g : geometric standard deviation	median: μ_g	$f(x, \mu_g, \sigma_g) = \frac{\exp\left(\frac{-(\ln x - \ln \mu_g)^2}{2 \ln^2 \sigma_g}\right)}{\sqrt{2\pi} \ln \sigma_g}$	$CV = \sqrt{\exp(\ln^2 \sigma_g) - 1}$
Normal	μ : arithmetical mean σ : arithmetical standard deviation	mean: μ	$f(x, \mu, \sigma) = \frac{\exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$	$CV = \frac{\sigma}{\mu}$
Uniform	a: minimum of the distribution b: maximum of the distribution	mean: $0.5(a+b)$	$\begin{cases} f(x, a, b) = \frac{1}{b-a} & \text{for } a < x < b \\ f(x, a, b) = 0 & \text{otherwise} \end{cases}$	$CV = \frac{b-a}{\sqrt{3}(b+a)}$
Triangular	a: minimum of the distribution b: maximum of the distribution c: most likely value of the distribution	most likely value: c	$\begin{cases} f(x, a, b, c) = \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a < x < c \\ f(x, a, b, c) = \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x < b \\ f(x, a, b, c) = 0 & \text{otherwise} \end{cases}$	$CV = \frac{\sqrt{a^2 + b^2 + c^2 - ab - ac - cb}}{\sqrt{2}(a+b+c)}$
Beta PERT	a: minimum of the distribution b: maximum of the distribution c: most likely value of the	Most likely value: c	$f(x, a, b) = \frac{(x-a)^{\alpha-1} (b-x)^{\beta-1}}{B(\alpha, \beta)(b-a)^{\alpha+\beta-1}}$	$CV = \frac{b-a}{a+4c+b}$

	distribution and $\alpha = 6 \frac{\mu - a}{b - a} \quad \text{and} \quad \beta = 6 \frac{b - \mu}{b - a}$ $\mu = \frac{a + 4c + b}{6}$			
Gamma	k : shape parameter λ : scale parameter	Most likely value: $\lambda(k-1)$	$f(x, k, \lambda) = \frac{x^{k-1} \exp(-x/\lambda)}{\Gamma(k) \lambda^k}$	$CV = \frac{1}{\sqrt{k}}$
Binomial	k : number of successes n : number of trials p : probability of success		$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$	$CV = \sqrt{\frac{1-p}{np}}$

However, the pedigree approach as applied in ecoinvent v2 had several weaknesses. One of them was the fact that the pedigree approach was only valid for lognormal distributions, while four different PDFs to model a datum's uncertainty were already available in the previous version of the ecoinvent database. While the lognormal distribution has many nice properties for modeling physical parameters (e.g. it is strictly positive, positively skewed and is the result of the multiplication of independent random variables (MacLeod et al. 2002; Slob 1994)), distributions other than the lognormal are more appropriate when they better represent the uncertainty associated with the datum. Most often, this will be the case when the basic uncertainty has been calculated based on available data. In these cases, the parameters of the basic uncertainty PDF will be directly known. In the cases where (1) the basic uncertainty cannot be calculated due to small sample size and (2) one nonetheless wishes to use a distribution other than the lognormal, the definition parameters of the PDF modelling a datum and its basic uncertainty can be determined using the default basic uncertainty factors used for the lognormal by converting the GSD^2 and the deterministic value into the definition parameters of the desired distribution.

The objective of this paper is to provide practical considerations on how to apply the pedigree approach for all distributions allowed in ecoinvent version 3. This specific objective treated in this paper belongs to a more global objective that is to improve the above approach and to legitimize the actual structure of uncertainty calculation in the ecoinvent database. The ecoinvent v3 database and its associated ecoSpold v2 data format will include three changes to the way uncertainty is handled:

1. Three new distributions will be available, the Gamma, the Beta (Pert approximation), and the Binomial distribution. The Binomial distribution is not affected by additional uncertainty, and is therefore not addressed further. The Erlang distribution is a special case of the Gamma distribution, and can therefore also be used. The definitions for these distributions are presented in Tableau 5-1. The uniform, triangular, normal and lognormal distributions were already available in ecoinvent v2.
2. The data quality indicator "sample size" and its corresponding additional uncertainty factors will be removed from the pedigree approach, as sample size should already be captured in the basic uncertainty factor. This should be recognized as a simplification, since the generic basic uncertainty factors do not capture the large deviations of small sample sizes often encountered in LCA.

3. The pedigree approach will be expanded to cover all six probability distribution functions.

This paper deals specifically with point 3. The methodology developed to combine basic and additional uncertainties for all PDFs and the resulting formulas are presented.

5.2.2 Materials and methods

5.2.2.1 Methodology applied to derive analytical formulas

In ecoinvent version 2, default basic and additional uncertainty factors were developed and a method to combine basic and additional uncertainty was proposed only for the lognormal distribution, used by default in the database. This combination is based on the fact that the spread of the distribution increases when adding an uncertainty component (the additional uncertainty here). In case of the lognormal distribution, the spread can be expressed through the 95th confidence interval $[\mu_g/\text{GSD}^{1.96}; \mu_g \cdot \text{GSD}^{1.96}]$. Conventionally, the value 1.96 is rounded to 2 and the 95th confidence interval is commonly expressed as $[\mu_g/\text{GSD}^2; \mu_g \cdot \text{GSD}^2]$. It has been shown that in case of experts' judgments and scarce data (characteristics that apply to the pedigree approach), the GSD^2 is a good measure of the spread of the distribution (MacLeod et al. 2002; Slob 1994). Because the basic uncertainty was expressed in terms of GSD^2 , the additional uncertainty factors were also expressed as “a contribution to the GSD^2 ” (Frischknecht et al. 2005).

Concretely, the combination that aims to correct the initial uncertain value using different components of variability, represented by the pedigree criteria, is done as follows:

A quantitative datum needed to compile a LCI is expressed with its basic uncertainty as a random variable (D_b). In the pedigree approach as implemented to date, D_b follows a lognormal distribution whose deterministic value is set as the median (μ_g) and the basic uncertainty factor defines the $\text{GSD}_{D_b}^2$. Then, the different components of the pedigree matrix act as modifiers of this random variable D_b , leading to the random variable D_T . This parameter is also a lognormally distributed random variable, and has the same deterministic value as D_b . However, the uncertainty factor, now represented by $\text{GSD}_{D_T}^2$, called the “total” uncertainty, represents the

initial basic uncertainty whose spread was increased to account for additional uncertainty. This additional uncertainty expresses the variability between a perfect case and the case described by the related pedigree matrix cell, and this for each pedigree criterion. For example, for the pedigree criterion “temporal correlation” and for a pedigree score equals to 3, the cell’s description is “Less than 10 years of difference to the time period of the dataset” (Weidema et al. 2013). For this specific case, the additional uncertainty must express the variation that exists in a 0 from 10 years time interval. In other words, the variation that occurs when compiling different data measured in this time interval. It can be defined as a random variable. The same can be said about the 5 pedigree criteria. D_i is also the random variable representing the additional uncertainty (for i from 1 to 5, representing the five pedigree criteria - from *reliability* to *further technological correlation*). These random variables were initially defined by experts’ judgments; they can also be determined using empirical sources as shown in the work by Citroth et al. (2013). Experts’ judgments are better represented using a lognormal distribution (MacLeod et al. 2002; Slob 1994); D_i are also represented as lognormal distributions with a geometric mean of one and with additional uncertainty factors as $GSD^2_{D_i}$. The lognormal distribution is also useful because, having geometric means of 1, D_i act as modifiers on D_b through $D_T = D_b \cdot \prod_{i=1}^5 D_i$ Équation 5-1 that results to an unmodified deterministic value. Moreover, the GSD^2 resulting from the multiplication of n independent lognormal distributions (D_b and D_i here) is expressed through

$$GSD^2_{D_T} = \exp\left(\sqrt{\ln^2 GSD^2_{D_b} + \sum_{i=1}^5 \ln^2 GSD^2_{D_i}}\right) \text{ Équation 5-2 (Limpert et al. 2001).}$$

As the pedigree criteria are independent (Weidema and Wesnæs 1996) so are D_T , D_b and D_i and Équation 5-2 can be used to combine basic and additional uncertainty leading to the formula applied in ecoinvent v2.2 (Frischknecht et al. 2005). Combination of random variables to express uncertainties can be found in environmental science (through assessment factors) and in some risk assessment methods (MacLeod et al. 2002; Vermeire et al. 1998; Slob 1994).

$$D_T = D_b \cdot \prod_{i=1}^5 D_i \text{ Équation 5-1}$$

$$GSD^2_{D_T} = \exp\left(\sqrt{\ln^2 GSD^2_{D_b} + \sum_{i=1}^5 \ln^2 GSD^2_{D_i}}\right) \text{ Équation 5-2}$$

The application of the pedigree approach to other PDFs will be directly based on how it was applied to the lognormal PDF and will follow the following four principles:

- *The additional uncertainty must modify neither the deterministic value (i.e. the statistical parameter used to define the exchange values in ecoinvent (Weidema 2013) nor the type of distribution chosen to represent the data (with its basic uncertainty).*
- *The total uncertainty is equal to the basic uncertainty when no additional uncertainty is added, i.e. when the data quality is assumed to be perfect and hence scores “1” for all data quality indicators using the pedigree matrix.*
- *The additional uncertainty is dimensionless and express the datum’s relative dispersion, according to each pedigree criteria, due to the use of imperfect data;*
- *The value of the additional uncertainty factors used for the lognormal distribution are used to derive the additional uncertainty for other PDFs. (See the Data Quality Guidelines of the ecoinvent database for the values of additional uncertainty factors (Weidema 2013)*

The lognormal and the normal distributions are intimately linked; if the random variable X is lognormally distributed, $\ln(X)$ is normally distributed and vice-versa. Équation 5-1 can also be translated into $\ln D_T = \ln D_b + \sum_{i=1}^5 \ln D_i$ Équation 5-3 where $\ln D_T$, $\ln D_b$ and $\ln D_i$ are normally distributed. That leads to Équation 5-4.

$$\ln D_T = \ln D_b + \sum_{i=1}^5 \ln D_i \quad \text{Équation 5-3}$$

$$\sigma_{\ln D_T} = \sqrt{\sigma_{\ln D_b}^2 + \sum_{i=1}^5 \sigma_{\ln D_i}^2} \quad \text{Équation 5-4}$$

So, if a datum D_b is originally normally distributed, Équation 5-4 can be expressed in terms of Équation 5-5 based, moreover, on the widely used formula to determine the standard deviation of a function of random independent variables.

$$\sigma_{D_T} = \sqrt{\sigma_{D_b}^2 + \sigma_{D_I}^2} \quad \text{Équation 5-6}$$

where D_I represents the random variable modelling the additional uncertainty that combines the five pedigree criteria.

The correspondences between the normal and the lognormal distribution can be found in Tableau 5-0-2 and Équation 5-7 can also be applied for the uniform, triangular, beta PERT and gamma distributions.

Tableau 5-0-2 Correspondences between the normal and the lognormal distributions

	Normal distribution	Lognormal distribution
Deterministic value	μ	μ_g
Dimensionless measure of variability	Coefficient of variation $CV = \frac{\sigma}{\mu}$	$\sigma_g = \exp(\sqrt{\ln(CV^2 + 1)})$ where σ_g is the GSD. $CV = \sqrt{\exp(\ln^2 \sigma_g) - 1}$ Alternatively
Confidence interval (68%)	$[\mu - \sigma; \mu + \sigma]$	$[\mu_g / \sigma_g; \mu_g \sigma_g]$

The standard deviation expresses an absolute dispersion around the mean; in the development of generic factors, that capture uncertainty information for all type of PDFs, the measure of dispersion should be expressed in terms relative to the median or mode for all the PDFs. This is done by using the relations between the the lognormal and the normal distributions. (see Tableau 5-0-2).

The coefficient of variation (CV) is chosen as a universal measure of variability. It is defined as the ratio between the arithmetic standard deviation and mean. More concretely, the CV measures a relative dispersion within a specific sample, regardless the distribution type considered. As it is a dimensionless measure of dispersion, the CV allows the comparison of the dispersion from different samples. It is also the initial measure of variation chosen to express the uncertainty with

the pedigree approach (Weidema and Wesnæs 1996). The total uncertainty will also be expressed as a CV. The uncertainty factors expressed as a GSD² applied in ecoinvent v2 can be easily translated into CV using equations presented in Tableau 5-0-2. In the same way, the expression of the additional uncertainty can easily be obtained using Eq. 2 and the definition of the CV using the GSD. The additional uncertainty expressed in term of CV (CV_I), is defined by Équation 5-8 where CV_i are the additional uncertainty factors linked to the five pedigree criteria.

$$CV_I = \sqrt{\prod_{i=1}^5 (CV_i^2 + 1) - 1} \quad \text{Équation 5-8}$$

The next step is also to express Équation 5-6 in terms of CV. Secondly; the standard deviation expresses an absolute dispersion around the arithmetic mean value. This dispersion - as well as the confidence interval around the mean- increases when additional uncertainty is added. In the case of symmetric distributions (such as the normal and uniform distributions), the mean is the mode. Since, for symmetric distribution, the arithmetic mean should not be modified when adding the additional uncertainty, the absolute dispersion is equivalent to the relative dispersion and Équation 5-9 can be used for symmetric distributions.

$$CV_T = \sqrt{CV_D^2 + CV_I^2} \quad \text{Équation 5-9}$$

In the case of asymmetric distributions, the arithmetic mean differs from the mode. For these distributions, the arithmetic mean will be affected by the consideration of additional uncertainty. While Équation 5-9 can still be used to calculate the relative dispersion parameter (CV_T), a new arithmetic mean (μ_T) that takes into account the effect of the additional uncertainty must be calculated using Équation 5-10 where μ is the arithmetic mean of the datum with its basic uncertainty.

$$\mu_T CV_T = \mu \sqrt{CV_D^2 + CV_I^2} \quad \text{Équation 5-10}$$

Having Équation 5-9 and Équation 5-10, the definition parameters of each distribution can be expressed through CV_T and also be defined for the total uncertainty.

5.2.2.2 Comparison between the analytical approach and a numerical approach

In order to test and validate the PDFs obtained through analytically derived formulas, they will be compared with that obtained using a numerical approach, the Monte Carlo simulation. The starting assumption that the additional uncertainty factors are described using GSD² remains unchanged, and distributions with total uncertainty are generated through a Monte Carlo simulation where the distribution modelling the data with its basic uncertainty is multiplied by the additional uncertainty represented by a lognormal distribution. The Monte Carlo simulation consists in randomly sampling the probability of each uncertain parameter (the data and its basic and additional uncertainty here) and then computing the sampled values using the model. The five distribution applied in ecoinvent v3 are tested, each with an arbitrary basic uncertainty factor (expressed as a CV), the same deterministic value and four different quintuplets giving a total of twenty tests. The simulation will be a 10 000 steps analysis performed by the ORACLE Crystal Ball release fusion edition (v 11.11.2.0) (ORACLE 2010). As the basic uncertainty and the additional uncertainty components are considered as independent (Weidema and Wesnæs 1996), covariances will not be considered in the simulation.

The Monte Carlo simulation is a good approximation that tends to theoretical when the number of steps increases. So, in order to compare the analytical and the numerical approaches for each distribution, relative errors will be calculated on the most relevant parameters representing the distribution and its total uncertainty: the standard deviation and the CV for the normal distribution; the minimum, the maximum and the CV for the uniform, triangular and beta PERT distributions; and the CV for the gamma distribution. In case of the distributions defined by location parameters (the uniform, triangular and beta PERT ones), the percentage of the resulting values obtained through the Monte Carlo simulation and that belongs to the interval defined by the minimum and the maximum of the distribution with its total uncertainty will also be determined.

5.2.3 Results and discussion

This section presents the obtained results for all distributions and the starting assumptions to derive the resulting formulas. Throughout, the subscript T will represent a distribution or parameter that models a datum with its total uncertainty.

5.2.3.1 Symmetric distributions

The normal and the uniform distributions are both symmetrical. Équation 5-9 is used, and the arithmetic mean is not modified by adding the additional uncertainty. Tableau 5-0-3 presents these assumptions in a mathematical form for both distributions and the resulting formulas to combine the basic and the additional uncertainties.

Tableau 5-0-3 : Assumptions and resulting formulas for symmetric distributions

Distribution	Assumptions	Formulas to combine basic and additional uncertainties
Normal	$\begin{cases} \mu = \mu_T \\ CV_T = \sigma_T / \mu \end{cases}$	$\begin{aligned} \mu &= \mu_T \\ \sigma_T &= \mu CV_T \end{aligned}$
Uniform	$\begin{cases} \mu = \frac{a+b}{2} = \frac{a_T+b_T}{2} = \mu_T \\ CV_T = \frac{b_T - a_T}{\sqrt{3}(a_T + b_T)} \end{cases}$	$\begin{aligned} a_T &= a + b - b_T \\ b_T &= \mu(1 + \sqrt{3}CV_T) \end{aligned}$

5.2.3.2 Asymmetric distributions

The triangular, Beta PERT and Gamma distributions are all asymmetric. The starting points to derive the formulas to combine basic and additional uncertainties are slightly different as the ones cited in the previous section. First of all, Équation 5-10 is used to determine CV_T and the deterministic value (the mode is not modified when the additional uncertainty is added.

Furthermore, the shape of the distribution is also not modified by adding the additional uncertainty. This last assumption is translated into a shape parameter (γ) that is defined for the triangular and the Beta PERT distribution. Tableau 5-4 presents these assumptions and the resulting formulas to combine the basic and the additional uncertainties.

A note can be added on the triangular distribution: the starting points lead to solve a quadratic equation. Although this quadratic equation has two solutions, only one is realistic here: the one that considers that b_T will be higher than b and a_T smaller than a (where b and a are respectively the maximum and the minimum of the probability density function).

Tableau 5-4 : Assumptions and resulting formulas for asymmetric distributions

Distribution	Assumptions	Formulas to combine basic and additional uncertainties
Triangular	$\mu = \frac{a+b+c}{3} \quad \text{and} \quad \mu_T = \frac{a_T + b_T + c}{3}$ $\left\{ \begin{array}{l} \gamma = \frac{c-a}{b-c} = \frac{c-a_T}{b_T-c} \\ CV_T^2 = 0.5 \frac{a_T^2 + b_T^2 + c^2 - a_T b_T - a_T c - b_T c}{(a_T + b_T + c)^2} \end{array} \right.$	$a_T = c(1+\gamma) - \gamma b_T$ $b_T = c + 3\mu \sqrt{CV_D^2 + CV_I^2} \sqrt{\frac{2}{1+\gamma+\gamma^2}}$
Beta PERT	$\mu = \frac{a+4c+b}{6} \quad \text{and} \quad \mu_T = \frac{a_T + 4c + b_T}{6}$ $\left\{ \begin{array}{l} \gamma = \frac{c-a}{b-c} = \frac{c-a_T}{b_T-c} \\ CV_T = \frac{b_T - a_T}{a_T + 4c + b_T} \end{array} \right.$	$a_T = c(1+\gamma) - \gamma b_T$ $b_T = c + \frac{\sqrt{CV_D^2 + CV_I^2}}{1+\gamma} (a + 4c + b)$
Gamma	$\left\{ \begin{array}{l} m = \lambda(k-1) = \lambda_T(k_T-1) \\ CV_T = 1/\sqrt{k_T} \\ \mu = k\lambda \quad \text{and} \quad \mu_T = k_T \lambda_T \end{array} \right.$ <p>where m is the most likely value and μ the arithmetic mean</p>	$k_T = 1 + \frac{m^2 + \sqrt{(2\mu m \sqrt{CV_D^2 + CV_I^2})^2 + m^4}}{4\mu^2 (CV_D^2 + CV_I^2)}$ $\lambda_T = \frac{m}{k_T - 1}$

5.2.3.3 Comparison with the results obtained through a Monte Carlo simulation

The total uncertainty resulting from the formulas available in Tableau 5-0-3 and Tableau 5-4 is compared with that obtained through a Monte Carlo simulation. The five new distributions with an arbitrary basic uncertainty factor, now expressed as a CV and the same most likely value (1.5),

are tested with four different pedigree quintuplets ((2;2;2;2;2), (3;3;3;3;3), (4;4;4;4;4), (5;5;5;5;5)), giving a total of twenty tests. The definition parameters for the PDFs without additional uncertainty, as well as those calculated for the total uncertainty are presented in Tableau 5-5. The graphics of the different resulting distributions are presented in Figure 5-1 (for the basic uncertainty and pedigree scores (3;3;3;3;3) and (5;5;5;5;5)) both for the analytical approach and distributions obtained through the Monte Carlo simulation (representing by the histograms on Figure 5-1). The relative errors on the relevant parameters are calculated. All the results from this comparison can be found in the Electronic Supplementary Material and only the major conclusions are presented here.

5.2.3.4 Application of the formulas

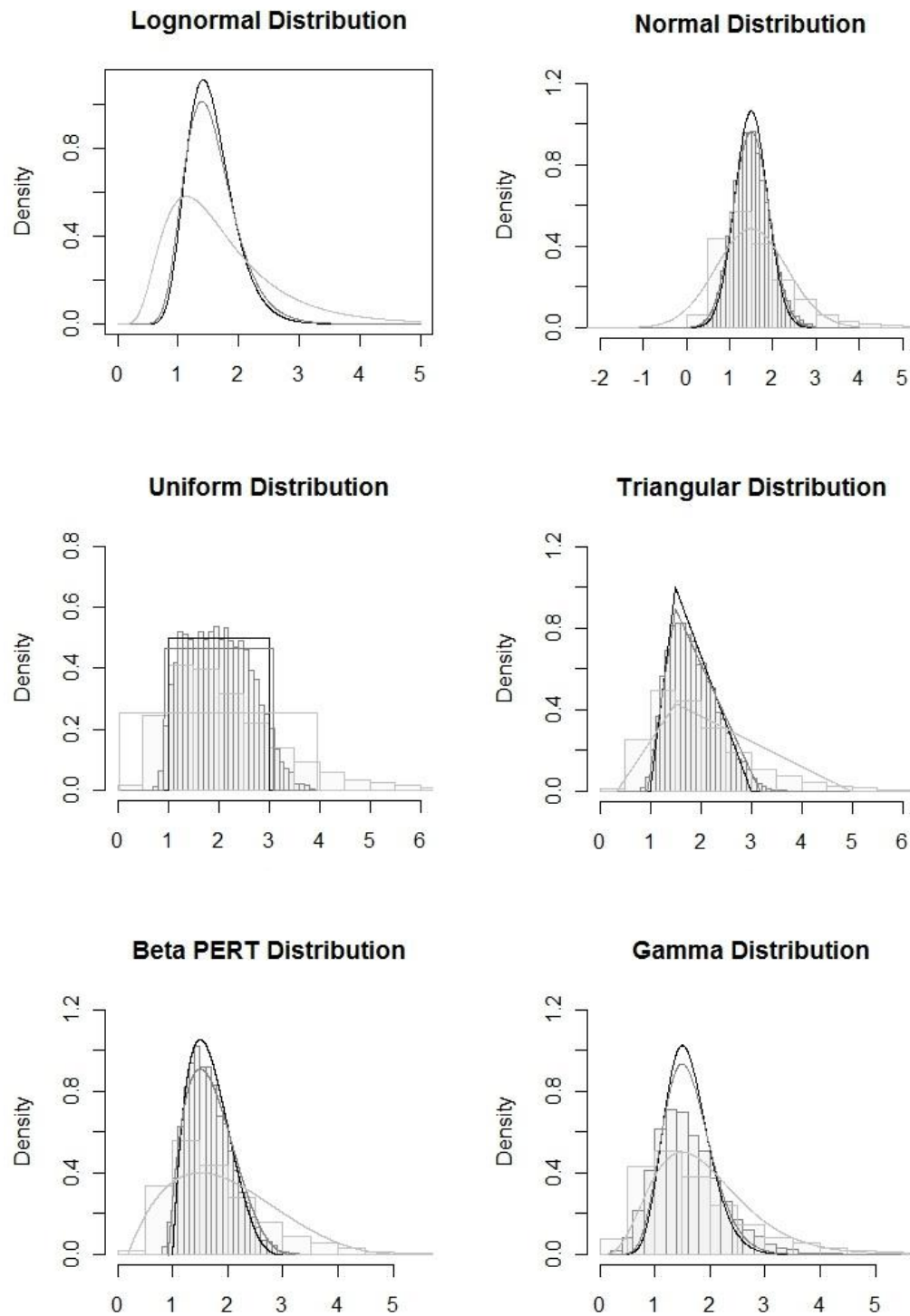
The proposed formulas can be applied as well to define the parameters of the desired distribution, keeping in mind that the used deterministic value should remain the same after adding additional uncertainty and should be the same as the one used as assumption in this proposed framework. Tableau 5-1 identifies the deterministic value for each distribution.

The formulas allow determining the CV of the data with its total uncertainty. If one chooses to express the uncertainty in a more usual fashion, for example with standard deviation σ_T , it can be retrieved from the following relation $\sigma_T = \mu_T CV_T$; where μ_T and CV_T are defined in Tableau 5-4.

Furthermore, these formulas, as well as the generic pedigree framework, should be applied keeping in mind that it is a semi-quantitative structure that aims to convert a codification of quality and lack of knowledge into uncertainty figures. If it is a good approximation when data characteristics are not known, the approach cannot be substituted by the use of raw data and statistics when these are available.

Tableau 5-5 : Definition parameters used to define the tested distribution that modelled a datum with its basic uncertainty and resulting calculated parameters when additional uncertainty is added. The basic and additional uncertainty combination is made using the proposed formulas foreseen in Tableau 5-0-3 and Tableau 5-4

PDF	Parameters	With basic uncertainty	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
Log normal	μ_g	1.5	1.5	1.5	1.5	1.5
	σ_g	1.279	1.289	1.313	1.416	1.690
Normal	μ	1.5	1.5	1.5	1.5	1.5
	σ	0.375	0.380	0.414	0.530	0.821
Gamma	k	16	15.66	13.47	8.92	4.72
	λ	0.1	0.102	0.120	0.189	0.403
Uniform	a	1	0.991	0.921	0.677	0.0386
	b	3	3.009	3.079	3.323	3.961
Triangular	a	1	0.993	0.940	0.765	0.336
	b	3	3.021	3.180	3.706	4.991
	c	1.5	1.5	1.5	1.5	1.5
Beta PERT	a	1	0.991	0.921	0.700	0.184
	b	3	3.028	3.237	3.901	5.450
	c	1.5	1.5	1.5	1.5	1.5



- Datum and its basic uncertainty
- Datum and its total uncertainty with pedigree scores (3,3,3,3,3)
- Datum and its total uncertainty with pedigree scores (5,5,5,5,5)

Figure 5-1 : Illustration of the different distributions foreseen in Tableau 5-5. The histograms represent the results obtained through the Monte Carlo simulation

5.2.3.5 Results and limitations

As seen for the Gamma distribution, the proposed approach has several limits. Some other limits coming from the applied methodology and the obtained results can be cited. First of all, although the used distributions served to model physical quantities, negative values can be obtained through the developed formulas (it happens in the given example: For the normal distribution when the additional uncertainty is determined through pedigree scores (5;5;5;5;5) the probability to obtain negative values is 3.10^{-2}). No sustainable way for avoiding this is proposed here; nevertheless three ways to consider negative values can be here cited:

- Consider them in uncertainty analysis, keeping in mind that they are fictive values;
- Define a threshold or a limit value (in this case 0) in the distribution definition. A certain percentage of possible values are, with this method, not considered;
- Transform all the negative values into a null value. The probability to get 0 will be nonetheless more important.

Another limit comes from the assumption that the additional uncertainty is best described using a lognormal distribution. This aspect will be checked in further work.

Another possible way to apply the pedigree approach to the other distributions will be to develop a computational solution based on Monte Carlo simulation. This would have some advantages. It would be possible to use different PDFs for different additional uncertainty types. For example, it is possible that, e.g. the basic uncertainty is best described by a lognormal distribution and the uncertainty associated with the temporal correlation (one of the indicator from the pedigree matrix) is best described by a triangular distribution. The here proposed approach do not permit to create a mix of distribution types as the additional uncertainty is assumed to be lognormally distributed. The computational approach based on Monte Carlo simulation will also dispense us of making some assumptions described in the previous section.

Last but not least, it should be noted that a mathematically more rigorous, but onerous, approach to determine the variance of a function of random variables is the convolution product. While

convolution product theoretically yields exact results, it is difficult to implement in software such as the ecoEditor and, depending on the assessed distribution, may actually lead to numerical approximations of the variance.

5.2.4 Conclusions

In order to apply the pedigree approach to all distributions foreseen in ecoinvent v3 it is first proposed to consider the CV as a dimensionless measure of variability. The proposed analytical formulas to combine basic and additional uncertainties for distributions other than lognormal are based on how this combination is made for the lognormal PDF and on the links between the multiplicative and the additive models. The different formulas are also based on the CV definition for each distribution. These obtained formulas fit, often quite well, with the combination of basic and additional uncertainties obtained through a Monte Carlo simulation. However, this proposed approach has some limitations. Some of them, particularly the starting assumption that the additional uncertainty is lognormally distributed, will be treated in future work. Indeed, despite its limits, based on the fact that it is a generic approach, the pedigree approach permits to consider the two elements that affect parameter uncertainty: data inaccuracy (through basic uncertainty) and lack of representative data (through additional uncertainty) in a semi-automatic way, explaining its use in a generic LCI database. See Huijbregts et al. for further details (2001)

These data variability assessment technics will be applied on all flow exchanges. They will also be applied to parameters which play an important role in ecoinvent v3 to calculate different exchange values. In order to be consistent in the whole ecoinvent database, the same approach – the pedigree approach – will be applied to define and quantify uncertainty on parameters.

CHAPITRE 6 ARTICLE 3: GIVING AN EMPIRICAL BASIS FOR UNCERTAINTY FACTORS USED IN GLOBAL LIFE CYCLE INVENTORY DATABASES: AN ALGORITHM TO UPDATE FACTORS USING NEW INFORMATION

6.1 Présentation de l'article

Ce troisième manuscrit présente les réponses au troisième objectif de cette thèse. Il a été **soumis** le 12 novembre 2015 à l'*International journal of life cycle assessment*. Le matériel supplémentaire soumis avec cet article se retrouve à l'Annexe D de ce document.

Les résultats présentés ici sont à la fois :

- Un développement méthodologique permettant la mise à jour de facteurs d'incertitude;
- La mise à jour des facteurs d'incertitude jusque-là utilisés dans la base de données ecoinvent; mise à jour qui s'appuie sur l'analyse de plus de 20 000 données d'ICV collectées.

6.2 Manuscrit

6.2.1 Introduction

6.2.1.1 Uncertainty modelling in the ecoinvent database

Even if life cycle assessment (LCA) results are often presented as single-point values, it is well known that these results are laden with uncertainty and variability. Uncertainty can be divided into three classes: model, scenario and parameter uncertainty (Lloyd & Ries, 2007; Reap et al., 2008). Whereas model and scenario uncertainties that arise during a LCA can be assessed through, for example, sensitivity analysis, parameter uncertainty can be treated through quantitative uncertainty assessment (by numerical methods as Monte Carlo simulation or by analytical methods) (Heijungs & Lenzen, 2014; Hong et al., 2010; Imbeault-Tétreault et al.,

2013). These techniques can however only be used if the uncertainty associated to the value of exchanges (elementary or intermediate flows) are quantified during the inventory stage of an LCA.

The ecoinvent database is the only LCI databases that systematically include uncertainty on data modeled in the database. It uses a semi-quantitative approach to estimate exchange value uncertainty. This approach is often referred to as the pedigree approach because it is based on the use of a pedigree matrix inspired by the NUSAP system (Funtowicz. & Ravetz 1990). It was originally developed for LCA by Weidema and Wesnæs (1996) and has been used in the ecoinvent database since 2005 (Frischknecht et al., 2005).

For a detailed description of the pedigree approach and the way it is applied in the ecoinvent database, the reader can refer to several papers (Ciroth et al., 2013; Frischknecht et al., 2005; Muller et al., 2014). In short, the approach ascribes a *basic uncertainty* value to exchanges, representing the intrinsic variability and stochastic error of the parameter, and increases this uncertainty using *additional uncertainty* factors that represent the use of imperfect data for the context of the study. By default, exchange values are assumed to be lognormally distributed, with the geometric mean defined as the deterministic value for the exchange and the geometric standard deviation estimated based on the basic and additional uncertainty factors that the pedigree approach provides. Distributions other than the lognormal can now also be used with the pedigree approach, see Muller et al. (2014).

If sufficient information can be accessed to calculate descriptive statistics and thus define the uncertainty, it should be used. If this is not the case, ecoinvent proposes default basic uncertainty values, expressed as GSD^2 and classified by type of exchange values and for three sectors: agriculture, combustion and process.

The additional uncertainty factors, also by default expressed as GSD^2 , are derived from a scoring of the data quality on five characteristics:

- *reliability of the data source*, scoring the quality of the sources and acquisition methods of the data used to quantify the exchange;
- *completeness*, scoring the statistical representativeness of the data;
- *temporal, geographical, further technological correlation*, scoring the degree to which the data used is representative of the time, area and technology of interest.

Using the cells descriptions of the pedigree matrix, scores of 1 to 5 (where 5 is the worst score) are given to each exchange value. The pedigree matrix in use in the ecoinvent v3 database is presented in Supporting information. These scores are then converted into so-called additional uncertainty factors, expressed as “contributors to the GSD²” (Frischknecht et al., 2005; Muller et al., 2014). The basic uncertainty and the additional uncertainty factors are then compiled, leading to a measure of the total uncertainty expressed as GSD².

6.2.1.2 The need to develop new uncertainty factors

This semi-quantitative pedigree approach is convenient to model the uncertainty on a large number of exchanges where real uncertainty information is not available. It however has several limitations. Three important limitations are (1) the imposition of the lognormal to describe the uncertainty of exchange values, (2) the reliance on experts’ judgments rather than on empirical data to quantify uncertainty factors; and (3) the use of additional uncertainty factors that ignore the type of exchange or industrial sector being assessed (Henriksson et al., 2014).

The first cited limitation was addressed by Muller et al. (2014), who extended the pedigree matrix approach to other types of distributions. The second limitation was partially addressed by Ciroth et al. (2013) who developed new additional uncertainty factors based on a statistical assessment of seven different data sources. The uncertainty factors they developed were slightly different from the ones used in ecoinvent v2 (A. Ciroth et al., 2013). However, they did not address the case of the basic uncertainty, they assumed that the factors are valid for all industrial sectors, and did not calculate factors for all “pedigree indicator-pedigree score” couples.

6.2.1.3 A way to update uncertainty information: the application of Bayes theorem

Published uncertainty factors associated with the pedigree matrix approach were either based on experts’ judgments (Frischknecht et al., 2005; Weidema & Wesnæs, 1996) or on the assessment of a limited number of data sources (Ciroth et al., 2013). While useful as first estimates of basic or additional uncertainty of exchanged, the empirical basis for these factors can be improved using more extensive data sources. A way to update quantitative information using new data

sources is to use Bayesian inference, coming from Bayes theorem as presented in Equation 6-1 (Qian et al., 2003).

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{\int_{\theta} p(d|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}} \quad \text{Equation 6-1}$$

with θ the random variable representing our variables of interest (in this case, the uncertainty factors) and \mathbf{d} the vector containing the new information. The elements of Equation 6-1 can be interpreted as follows:

- $p(d|\theta)$ is the likelihood function : it describes the assumption that the data \mathbf{d} were observed based on θ ;
- $p(\theta)$ is the prior distribution : it represents the knowledge (often subjective) that is available on θ ;
- $p(\theta|d)$ is the posterior distribution : it represents all the information we finally have on θ . The distribution's mean can become an estimate for θ and its confidence interval an estimate on θ 's uncertainty;
- The denominator can be considered as a standardization coefficient allowing the posterior probability distribution to take its value in the interval [0,1].

While the use of Bayesian inference in LCA is mentioned in several papers (Björklund, 2002; Huijbregts, 1998; Katz, 2002), only few describe concrete applications. Ukidwe et al. (2004) and Miller et al. (2013) used the Bayesian approach to refine their LCI models; Lo et al. (2005) used it to better estimate the uncertainty of their systems by first identifying the greater contributors to the uncertainty and then applying the Bayes theorem to these greater contributors to refine their associated uncertainty.

The aim of this paper is to strengthen the pedigree approach by refining existing uncertainty factors through the consideration of new data sources, developing uncertainty factors that are specific to industrial sectors, and defining new basic uncertainty factors. In order to do so, Bayesian inference will be used to both combine information already known – by experts'

judgments and data assessment – with new information obtained through empirical data assessment.

6.2.2 Methods

6.2.2.1 General methodology

Tableau 6-1 is a summary of the three-step methodology developed to obtain updated basic and additional uncertainty factors. These three steps are detailed in the following paragraphs.

First of all, a descriptor of the uncertainty factors should be chosen. Weidema and Wesnaes (1996) used the coefficient of variation (CV, ratio between the standard deviation and the mean) to describe the uncertainty factors. In the ecoinvent v2 database, the GSD^2 is used as a descriptor of the uncertainty factors, more specific the uncertainty factors are described as “contributors to the GSD^2 ” (Frischknecht et al., 2005). Then, in the third version of the ecoinvent database, they were expressed as “square of the standard deviation of the underlying normal distribution”, i.e. the square of the standard deviations of the variable’s logarithm (Weidema et al., 2013). In this paper, to remain consistent with the previous work done by Ciroth et al. (2013), the GSD^2 is chosen as a descriptor of the uncertainty factors. The GSD^2 can easily be transformed in the descriptor used in ecoinvent v3 (σ^2) using $\sigma^2 = (\ln(GSD))^2$ Equation 6-2.

$$\sigma^2 = (\ln(GSD))^2 \quad \text{Equation 6-2}$$

Tableau 6-1 The three steps methodology used to develop updated uncertainty information

1. Identify initial estimate of uncertainty factors	1.1 Census of formerly published uncertainty figures 1.2 Translation of these figures into uncertainty factors
2. Collect new information	2.1 Data harvesting and preparation <ul style="list-style-type: none"> - Data source selection - Data harvesting - Data classification into subgroups that isolate basic uncertainty and the different components of additional uncertainty 2.2 Estimation of empirical uncertainty factors <ul style="list-style-type: none"> - GSD² calculations for each pedigree indicator, for specific industrial sectors and for the subgroups created for the basic uncertainty - GSD² transformation to “contributors to the GSD²” (i.e. uncertainty factors)
3. Update uncertainty information	3.1 Estimation of the <i>prior distribution</i> based on the formerly published uncertainty factors 3.2 Estimation of the <i>likelihood function</i> based on the obtained empirical uncertainty factors 3.3 Application of Bayes theorem to obtain new uncertainty factors that combine expert judgements and empirical data

6.2.2.2 Identify initial estimate of uncertainty factors

The basic and additional uncertainty factors used in ecoinvent v2 and v3 and as published in ecoinvent v3 data quality guidelines (Weidema et al., 2013) are coming from experts' judgments and expressed as the “square of the standard deviation of the underlying normal distribution” (σ^2).

To express as “contributors to the GSD^2 ”, the reverse transformation of the one performed between ecoinvent v2 and ecoinvent v3 can be applied. Equation 6-3 can so be used and the uncertainty factors (UF) can then be expressed by Equation 6-4.

$$GSD = \exp(\sigma) \quad \text{Equation 6-3}$$

$$UF = (\exp(\sigma))^2 \quad \text{Equation 6-4}$$

*where UF is the
value of the
uncertainty factor*

When the pedigree matrix was first developed in 1996, Weidema and Wesnaes also proposed uncertainty figures for the additional uncertainty based on estimates and expressed as CVs. These estimates are also used as prior information once translated into GSD^2 . In order to perform this translation $GSD = \exp\left(\sqrt{\ln(CV^2 + 1)}\right)$ Equation 6-5 can be used. This use is based on the following:

- Uncertainty factors are expressed as contributors to the GSD^2 ;
- The pedigree approach was developed for lognormally distributed exchange values;
- Equation 5 links the GSD and the CV for a lognormal distribution.

$$GSD = \exp\left(\sqrt{\ln(CV^2 + 1)}\right) \quad \text{Equation 6-5}$$

Finally, the values developed by Ciroth et al. for the additional uncertainty, both in a published paper (2013) and in a report delivered to the ecoinvent centre (2012), will also be used as prior information for this work.

These four data sources for prior information lead to 1 datum for each component of the basic uncertainty and to 2 to 26 for each component of the additional uncertainty. These data can be found in the Supporting information. In order to remain consistent, one outlier from the work of Ciroth et al. (2012) was removed from the analysis.

6.2.2.3 Collecting new information

1. **Data harvesting and preparation**

The selection of the data sources used in this assessment is based on the following points:

- The typologies of data that arise in LCI must be considered: the harvested data must represent both intermediate and elementary flows (together referred to as exchanges);
- It must be possible to normalize data to a reference flow, i.e. a unit output from an activity;
- The data sources must cover a large spectrum of industrial sectors, years, geographical areas and types of sources to have a sufficient set of data for each assessed “pedigree indicator-pedigree score” couple.

Based on these points, data from publicly available LCA reports, published LCA papers, emissions factors databases and sector specific LCI databases were collected. The 68 different data sources used, listed in the supporting information, yielded 23,200 data points for analysis.

These data are compiled in a single database that contains all the information needed to perform the assessment:

- The type and the name of the datum, its value and its corresponding unit;
- The reference flow to which the datum is normalized, its value and its corresponding unit;
- The industrial sector from which the data was generated. The industrial sectors classification used is based on the 2012 North American Industry Classification System (NAICS) and on the 1997 Selected Nomenclature for sources of Air Pollution (SNAP97) for the combustion sector.
- The other information useful to classify the data into a specific “pedigree indicator-pedigree sector” couple: the year and the geographic area where the data were generated and how it was generated (e.g. estimated or calculated).

The compiled data have, at this point, different units and are linked to different reference flows. All data were converted to SI units and normalized to a unit amount of reference flow.

Finally, this compilation into a single database allows the classification of the different data into specific subsets for the basic uncertainty and the different “pedigree indicator-pedigree score” couples for different industrial sector (based on the NAICS – level2) for the additional uncertainty. The basic uncertainty is defined as the uncertainty that remains when all pedigree scores equal 1; in each created subgroup, the data are coming from a same technology, a same year and a same country and their reliability score equals 1. The data classification for the additional uncertainty is based on the pedigree matrix cells description (as described in Weidema et al. (2013)). The way the pedigree matrix’s cells were interpreted to create the different subgroups is available in the Supporting information section.

In practical considerations, for all pedigree indicators (except for the completeness indicator which case is described in the paragraph 2.3.3), the subgroups creation are based on the principle of “Russian dolls” as described in Figure 6-1.

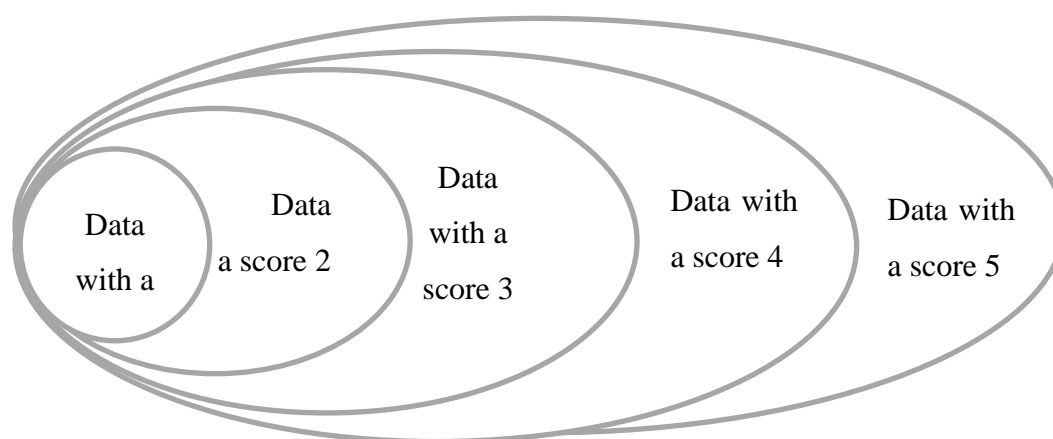


Figure 6-1 Illustration of the « Russian dolls » principle used to classify data into subgroups for the additional uncertainty

EXAMPLE 1: DATA CLASSIFICATION BY “PEDIGREE INDICATOR – PEDIGREE SCORE” - Case of the temporal correlation indicator

Using a reference year of 2013 and the pedigree matrix cell description given in the Supporting information, all data points can be associated with a score based on its age. For the score 1, associated with no additional uncertainty, all data harvested coming from the years 2013 to 2010 are kept. The other subgroups are then built following the “Russian dolls” principle illustrated in Figure 6-1, all the data from 2013 to 2007 are kept to calculate score 2; the data from 2013 to 2001 for score 3; all the data from 2013 to 1998 for score 4; and, finally, all the data harvested for score 5.

2. Estimation of empirical uncertainty factors

The estimation of empirical uncertainty factors are based on the formulas proposed by Citroth et al. (2013). In a first place, the GSD^2 for each created subgroup is calculated based on Equation 6-6. For the additional uncertainty, these GSD^2 must be converted into uncertainty factors using Equation 6-7. The basic uncertainty factor is directly defined as the calculated GSD^2 .

$$GSD = \exp \left(\sqrt{\frac{1}{n} \sum_{i=1}^n \ln \left(\frac{x_i}{\bar{x}_g} \right)^2} \right)$$

Equation 6-6

where n is the number of data in the sample and \bar{x}_g the geometric mean of the sample

$$UF_i = \begin{cases} \frac{GSD_{i,j}^2}{GSD_{i,1}^2} & \text{if } GSD_{i,1}^2 \leq GSD_{i,j}^2 \\ \frac{GSD_{i,1}^2}{GSD_{i,j}^2} & \text{else} \end{cases}$$

Equation 6-7

where UF_i is the uncertainty factor for the i^{th} pedigree indicator;

$GSD_{i,1}^2$ is the geometric standard deviation for the i^{th} pedigree indicator and the score 1;

$GSD_{i,j}^2$ is the geometric standard deviation for the i^{th} pedigree indicator and a score j .

EXAMPLE 2: HOW THE UNCERTAINTY FACTORS ARE CALCULATED - Case of the temporal correlation indicator

Following the EXAMPLE 1 above, subgroups are created based on the industrial sector (based on 2-digit NAICS codes) and the type of exchange (e.g. specific emission, specific type of intermediary input, etc.). For example, all the CO₂ emissions for the primary metal manufacturing sector were classified into a specific subgroup.

In this subgroup, the GSD² of all data having a pedigree score of 1 is calculated. The same is done for the data having scores from 2 to 5. Then, Equation 6-7 is used to calculate the uncertainty factors for CO₂ emissions of the primary metal manufacturing sector.

3. The completeness indicator case

In Bo Pedersen Weidema and Wesnæs (1996), the completeness indicator expresses the variability due to “the number of data collection points, the period of collection and the representativeness to the total population”; in ecoinvent v3, the completeness indicator expresses the representativeness of the data regarding the relevant sites of the considered market (B.P. ; Weidema et al., 2013). This indicator can also be directly linked to the representativeness of the used sample to describe the entire population (and so here, to produce a representative LCI datum). The question of the effect of sampling on the variability can be answered statistically by building a confidence interval for the population variance knowing the size of the of the sample regarding the whole population (as it is done for opinion surveys for example).

Confidence interval estimation for a population is well known when the population is assumed to be normally distributed. For normally distributed data, the ratio $n.s^2/\sigma^2$ (where n is the size of the population, s the sample standard deviation and σ the standard deviation of the population) follow a chi-square distribution with $n-1$ degree of freedom. For a 95% confidence level of, the confidence interval estimation for the population variance can be obtained through Equation 6-8.

$$\frac{ns^2}{\chi_{\alpha_1}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\alpha_1}^2}$$

Equation 6-8

As before, Equation 6-3 can be used to transform σ^2 into GSD^2 . Equation 6-7 is then used to derive uncertainty factors for the completeness indicator. In Equation 6-8, $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ can be found in the table of the χ^2 distribution; $\alpha/2$ is linked to the confidence level that need to be reached, for a symmetric confidence interval estimation and a confidence level of 95%, $\alpha/2$ is then equal to 0.025.

Concerning the value of s , it can be directly linked to the basic uncertainty as it represents the value of the sample standard deviation that can be calculated by the person that defines the LCI datum with its uncertainty. For the rest of the paper, it will be assumed that s^2 equals 1. It must be here underlined that the value can be changed by the user if necessary.

The factors developed through this method will not directly fit the pedigree matrix cell descriptions for the completeness indicator but the ones indicating in Tableau 6-2. Supposing that the whole population is represented by 100 sites, the values of n are respectively: 100, 75, 50, 25 and 5 for the five pedigree scores. Concerning the fifth score and its cell matrix description; the choice of $n=5$ is arbitrary. If the representativeness is unknown ($n=0$), in theory the associate uncertainty will be infinite. This choice is made here to avoid this situation.

Tableau 6-2 Proposed pedigree matrix cell description for the completeness indicator

1	2	3	4	5
Representative data from all sites relevant from the market considered.	Representative data from >75% of the sites relevant from the market considered.	Representative data from >50% of the sites relevant from the market considered.	Representative data from >25% of the sites relevant from the market considered.	Representative data from less than 5% of the sites relevant from the market considered.

6.2.2.4 Updating uncertainty information

The collection of new information doesn't permit to have data that represent all industrial sectors and all type of technologies and that for all type of exchanges that arise in the LCI phase. Rather than using only partial information to define new uncertainty factors, the initial estimates of

uncertainty factors will be updated using the new, quantitative but partial, information and the Bayes theorem as described in Equation 6-1. Before applying the theorem, the parameter of interest (i.e. the information that will be refined thanks to the theorem) needs to be defined. Here this parameter is the uncertainty factor (UF). Then, and in order to apply the Bayes theorem for basic uncertainty and each “pedigree indicator-pedigree score” couple, three more components are needed:

- The prior distribution of UF : UF needs to be defined as a random variable whose prior distribution represents all the information we have on the uncertainty factor. Defining UF as a random variable means that variability on the uncertainty factor itself is introduced;
- The data d : d is here the vector containing the new empirical estimation of the uncertainty factors based on the newly assessed data sources;
- The likelihood function f : f links the data to the parameter of interest. It is here assumed that the data can be described by $f(UF)$ where the random variable UF is the mean of the likelihood function.

The application of the Bayes theorem depends on the nature of the prior distribution and the likelihood function. If the numerator in Equation 6-1 can be calculated “easily”, it’s not the same with the denominator that needs to be determined numerically using Bayesian Monte Carlo or Markov chain Monte Carlo when the prior distribution and the likelihood function are not conjugate (Ben Letham, 2012; Qian et al., 2003).

In this application of the Bayes theorem, the information we have on the prior distribution of the uncertainty factors is very limited. In order to define the prior distribution, and also the likelihood function, the distribution of the newly obtained uncertainty factors is drawn. When representing the distribution of the new obtained uncertainty factors for the basic uncertainty and all “pedigree indicator – pedigree score” couples, two tendencies can be noted:

- All the factors are positive (by definition);
- For a majority of the factors, their distributions are right-tailed.

Given these two characteristics, the lognormal distribution is chosen to represent both the prior distribution and the likelihood function. This assumption of lognormally distributed uncertainty factors is tested on the new empirical uncertainty factors (see Supporting information). A

lognormal distribution is not directly conjugate to another lognormal distribution. However, two normal distributions are, and they result in a normally distributed posterior distribution. In order to perform the Bayesian application, we use the fact that, if X is a lognormally distributed random variable with parameters μ and σ , then $\ln(X)$ is normally distributed with the same parameters

(see $LN(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$ Equation 6-9 and and

$N(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$ Equation 6-10). Figure 6-2 makes a census of all the

assumptions needed to derive the prior distribution and the likelihood function, as well as the calculation performed to obtain the posterior distribution. The most subjective assumptions are made on the parameters of the prior distribution due to the lack of available information. In order to derive these parameters, the mean of the prior distribution is set as the value of the uncertainty factor as found in ecoinvent v3 and the standard deviation is derived using the 95th interval if sufficient data are available (see Figure 6-2).

$$LN(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \text{ Equation 6-9}$$

and $N(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$ Equation 6-10

where $LN(x, \mu, \sigma)$ stands for the lognormal PDF and $N(x, \mu, \sigma)$ for the normal PDF.

Once the parameters of the posterior distribution have been derived, it is possible to determine the updated uncertainty factors. The updated uncertainty factor is defined as the mean of the posterior distribution that can be defined using Equation 6-11 (which is the equation linking the mean of a lognormal distribution to the logarithmic parameters that defines the distribution).

$$UF_p = \exp(\mu_p + 0.5\sigma_p^2) \text{ Equation 6-11}$$

EXAMPLE 3: HOW THE INFORMATION IS UPDATED - Case of the temporal correlation indicator

Take the temporal correlation indicator for a score 2. The uncertainty factor for this case is noted UF_{TC2} . To update the information, the prior distribution of UF_{TC2} and the likelihood function need to be defined.

For the prior distribution, all the uncertainty factors published in previous paper are used to define the distribution based on the assumptions presented in Figure 1. The likelihood function is obtained by calculating the logarithmic mean and the logarithmic standard deviation of all the uncertainty factors calculating by subgroups for a score 2 (uncertainty factors obtained in the case of the CO₂ emissions for the primary metal manufacturing sector is among them). These logarithmic mean and logarithmic standard deviation define the lognormal distributions used as likelihood function.

6.2.3 Results

6.2.3.1 The case of the additional uncertainty factors

The results obtained for the additional uncertainty factors are presented in Figure 6-3 and Tableau 6-3. Figure 6-3 shows the values of the generic factor for all “pedigree indicator-pedigree score” couples. These generic values can also be found in Tableau 6-3 that adds the additional uncertainty factors obtained by industrial sectors. Globally, the likelihood values are greater than the prior values. That can be explained by the large variability present in the subgroups of assessed data. Even if these data are grouped according to specific industrial sectors, these sectors remain large (for example, the manufacturing sector contains both cement production and pulp and paper production) explaining large variability and large additional uncertainty factors especially for pedigree scores equal to 5. The fact that the posterior values lay between the prior and the likelihood can be explained by the application of the Bayes theorem that gives a weighted compromise between the prior knowledge and the data.

Some of the calculated uncertainty factors are counter-intuitive. First, for some pedigree indicators, the likelihood value is the same for two consecutive pedigree scores (see for example the scores 4 and 5 for the geographical correlation indicator in the manufacturing sectors). These cases arise when the same data constitute the subgroup for each score, due to a lack of data. One exception is the equality of scores 4 and 5 for the generic factor in the further technological indicator, which is not a real equality and comes from the expression of the uncertainty factor using only three digits. Second, some uncertainty factors are greater for a score 4 than for a score 5 (see for example the temporal correlation indicator for the transportation sector or the geographical correlation indicator for the combustion and agricultural sectors. This arises when a subgroup with more data (for a score 5) is less variable than the subgroup with a score 4.

One should also note that the results are sensitive to the assumption made on the prior standard deviation (see Figure 6-3). The effect on the assumption for the additional uncertainty factors expressed by industrial sector can be found in the Supporting information.

The values calculated using the Bayesian approach are more complete than those calculated by Ciroth et al. (2013), i.e. there are uncertainty factors for each “pedigree indicator-pedigree score” couple. They are also generally greater for the higher scores.

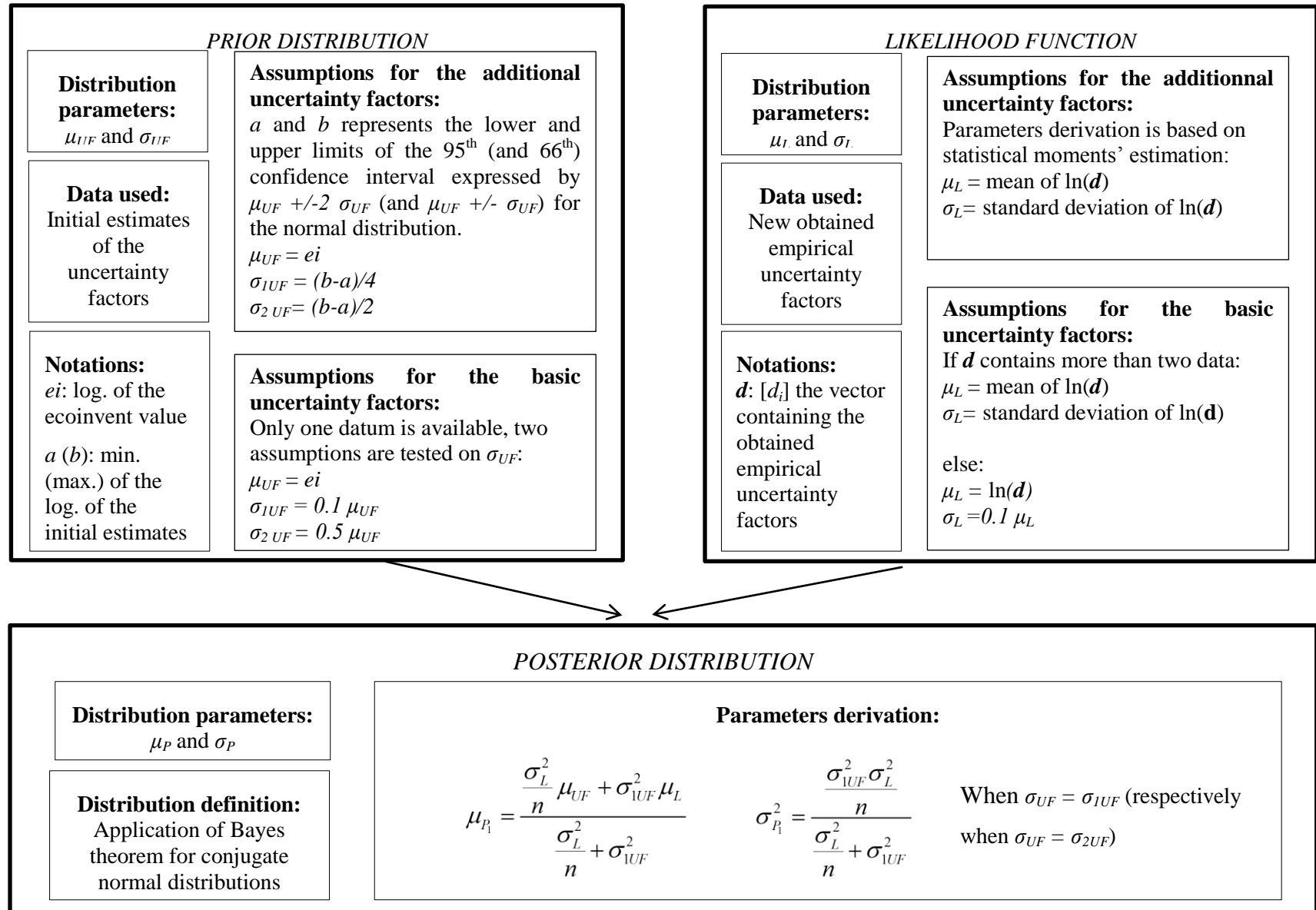


Figure 6-2 Assumptions used to apply the Bayes theorem – Notations log. (logarithm), min. (minimum), max. (maximum)

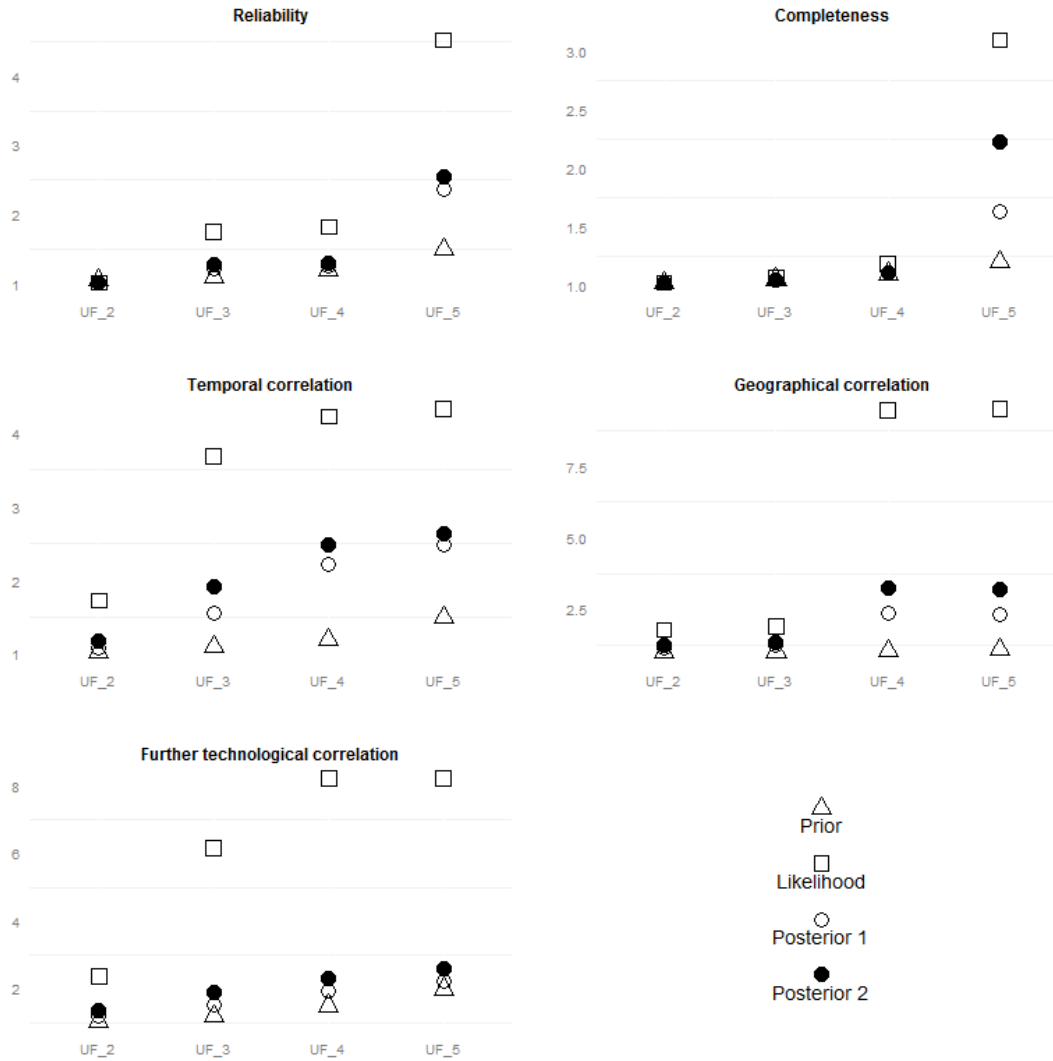


Figure 6-3 Additional uncertainty factors representation for the 5 pedigree indicators and the 4 pedigree scores. The value of Posterior 1 is obtained using σ_1 for the prior distribution, the value of Posterior 2 is obtained using σ_2 (see Figure 6-2)

Tableau 6-3 Updated additional uncertainty factors obtained by using σ_{UF1} for the prior distribution (see Figure 6-2), the results for σ_{UF2} can be found in the Supporting information. The prior is theecoinvent value; L stands for the likelihood value and P for the posterior value.

Pedigree indicator		Prior	Generic		Agriculture		Combustion		Utilities		Manufacturing (other)		Chemical products manufacturing		Metal manufacturing		Transportation	
			L	P	L	P	L	P	L	P	L	P	L	P	L	P	L	P
Reliability	UF2	1.05	1.01	1.01	1.00	1.00	1.06	1.06			1.00	1.00			1.00	1.00		
	UF3	1.10	1.75	1.21	1.05	1.05	1.18	1.12			2.26	1.20			1.00	1.00		
	UF4	1.20	1.82	1.25	1.05	1.07	1.18	1.18			2.26	1.25			1.40	1.21		
	UF5	1.50	4.52	2.36	1.98	1.60	4.61	1.69			6.20	2.60			1.64	1.51		
Completeness	UF2	1.02	1.02	1.02														
	UF3	1.05	1.07	1.05														
	UF4	1.10	1.18	1.10														
	UF5	1.20	3.09	1.63														
Temporal correlation	UF2	1.03	1.73	1.08	3.41	1.03	1.47	1.08	1.00	1.00	3.58	1.05			1.18	1.07	1.20	1.16
	UF3	1.10	3.69	1.55	3.41	1.13	2.33	1.27	5.66	1.22	7.60	1.38			1.44	1.22	1.32	1.26
	UF4	1.20	4.23	2.22	5.03	1.80	2.68	1.72	9.52	1.87	6.34	2.12			1.44	1.32	1.27	1.26
	UF5	1.50	4.33	2.49	5.03	4.07	3.12	1.75	9.53	2.52	6.12	2.47			1.58	1.49	1.15	1.15
Geographical correlation	UF2	1.01	1.78	1.14	1.00	1.00	1.29	1.08			3.12	1.09						
	UF3	1.02	1.90	1.23	1.72	1.10	1.29	1.11			3.12	1.16						
	UF4	1.05	9.45	2.36	4.14	1.57	4.73	1.66			23.60	1.98						
	UF5	1.10	9.49	2.30	4.38	1.55	4.73	1.65			23.60	1.92						
Further technological correlation	UF2	1.05	2.37	1.19	1.00	1.00	1.04	1.04	1.00	1.00			8.58	1.16				
	UF3	1.20	6.17	1.52	3.33	1.32	1.03	1.04	4.62	1.29			14.47	1.44				
	UF4	1.50	8.23	1.95	3.73	1.59	1.54	1.50	8.13	1.62			14.69	2.04				
	UF5	2.00	8.23	2.23	3.72	2.03	1.54	1.89	3.03	2.00			35.02	2.21				

6.2.3.2 The case of the basic uncertainty factors

The subgroups created for the basic uncertainty factors leads to the development of fewer factors than the ones originally present in the basic uncertainty table in the ecoinvent database. The results can be found in Tableau 6-5. The obtained likelihood values and the uncertainty values for all obtained subgroups are available in the Supporting information. As for the additional uncertainty factors, the posterior values are sensitive to the assumptions made on the standard deviation of the prior distributions.

While for most values, the posterior values are only slightly different from their priors, differences are more marked for the particulates emissions (PM10 and PM2.5). This can be explained by the large number of subgroups (respectively 9 and 19 on 64 subgroups assessed to derive basic uncertainty factors) were used to determine the likelihood function, subgroups that represent different type of technologies (from battery manufacturing to steel foundries, see Supporting information).

6.2.3.3 What if new data are available?

Using Bayesian inference allows updating the uncertainty factors when new data are available. Three different cases for the update can be cited. All of them use the posterior values found in the previous section and the posterior distributions that are described in the Supporting information. For an update, the posterior distribution becomes the new prior distributions.

CASE 1: Data are available to update generic factors; the data are used to define the likelihood function that, combined to the new generic prior distribution, leads to an updated generic posterior distribution of the uncertainty factor.

CASE 2: Data are available for a specific industrial sector i for which a new prior distribution is available. The data are used to define the likelihood function for the sector i that, combined to the prior distribution for the sector i , lead to an update posterior distribution of the uncertainty factor for the specific industrial sector i .

CASE 3: Data are available for a specific industrial sector j for which there is no specific new prior distribution available. The data are used to define the likelihood function for the sector j ,

and this likelihood function is combined to the generic prior distribution, leading to an updated posterior distribution for the specific industrial sector j . These data can also be added to the original data (as the sector j was not represented): this new group of data allows one to define a new generic likelihood function that, combined to the generic prior distribution, leads to an updated generic posterior distribution of the uncertainty factor.

The following example illustrates a *CASE 3* situation. Using the data of the INIES database (a French database for the residential sector (INIES, 2013)) and, more specifically, the data regarding the manufacturing of boards for walls and ceiling, a likelihood function for the manufacturing sector and the further technological indicator can be defined. The data used to derive the likelihood function can be found in the Supporting information. The new obtained uncertainty factors for the manufacturing sector can be found in Tableau 6-4 and is illustrated in Figure 6-4 for the pedigree score 4.

Tableau 6-4 Updated additional uncertainty factors for the further technological correlation indicator for the manufacturing sector using a new data source

	UF_2	UF_3	UF_4	UF_5
Prior value	1.19	1.52	1.95	2.23
Likelihood value	1.95	2.22	2.40	NA
Posterior value	1.26	1.69	2.07	NA

Tableau 6-5 Obtained posterior values for the basic uncertainty factors. The posterior 1 values were obtained using σ_{UF1} (see Figure 6-2) and the posterior 2 values using σ_{UF2}

	Combustion			Process			Agriculture		
	Prior	Posterior 1	Posterior 2	Prior	Posterior 1	Posterior 2	Prior	Posterior 1	Posterior 2
Thermal energy, electricity, semi-finished products, material, waste	1.05	1.05	1.11	1.05	1.05	1.06	1.05	1.05	1.05
Transport	2.00			2.00			2.00		
Infrastructure	3.00			3.00			3.00		
Primary energy carriers, metals, salt	1.05			1.05			1.05		
land use occupation	1.50			1.50			1.10		
land use transformation	2.00			2.00			1.20		
WATER									
BOD, COD, TOC, DOC, inorganic compounds				1.50	2.13	2.13			
Individual hydrocarbons, PAH				3.00	3.25	3.57			
Heavy metals				5.00			1.80		
Pesticides							1.50		
NO ₃ , PO ₄							1.50		
land use occupation									
land use transformation									
SOIL									
Oil, hydrocarbons				1.50					
Heavy metals				1.50			1.50		
Pesticides							1.50		
AIR									
Carbon dioxide	1.05	1.06	1.15	1.05					
SO ₂	1.05								
NM VOC total	1.50								
Nox, N ₂ O	1.50						1.40		
CH ₄ , NH ₃	1.50						1.20	1.20	1.26
Individuals hydrocarbons	1.50			2.00					
PM _{>10}	1.50			1.50					
PM ₁₀	2.00			2.00	2.12	5.46			
PM _{2.5}	3.00			3.00	3.74	18.74			
PAH	3.00								
CO, Heavy metals	5.00								
Inorganic emissions				1.50	1.56	3.32			
Radionuclides				3.00					

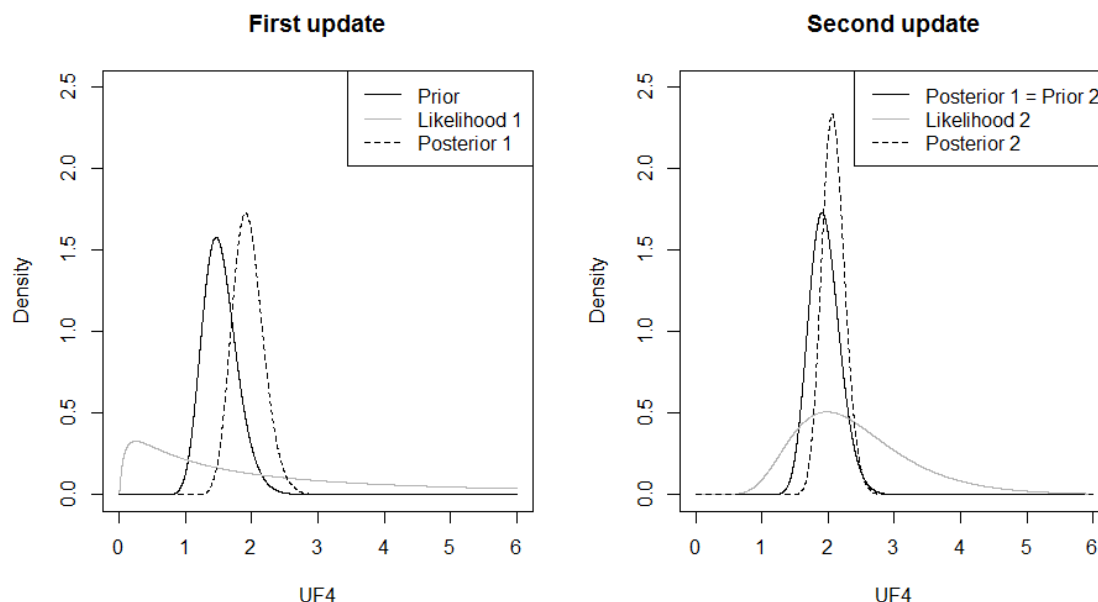


Figure 6-4 Representation of the further technological correlation indicator for the manufacturing sector updated two times: once using the the master datatable used in this study, the second time by accessing a new data source from the INIES database

6.2.4 Discussion and conclusion

6.2.4.1 Limitations due to the methodology

The results described in the preceding section permit to have empirical additional uncertainty factors for some specific sectors and new empirically based basic uncertainty factors. These results rely on some assumptions, some of which are directly linked to the application of the Bayes theorem.

A first assumption is the definition of the fifth pedigree score for the completeness indicator (see section 6.2.2.3). In order to assess the effect of this choice, the case where $n=1$ was assessed leading to an uncertainty factor greater than 3.10^9 . To avoid this situation, the value $n=5$ is here kept.

Another assumption is the choice of the logarithm distribution as the distribution for modelling the uncertainty factors (for both the prior distribution and the likelihood distribution). This choice

is only based on data representation and was tested (the assumptions of lognormally distributed uncertainty factors is at 95% true for one third of the assessed uncertainty factors, see Supporting information). This choice is also made to simplify the use of the Bayes theorem by using conjugate distributions. Nonetheless, the specific distribution of the uncertainty factors is not, per se, the parameter of interest here; the mean of the posterior distribution is here the parameter used to describe the uncertainty factors.

The definition of the distribution parameters, especially for the prior distribution for the basic uncertainty factors, are also based on assumptions that affect the results, since the standard deviation of the prior distribution is used to derive the mean of the posterior distribution (see Figure 6-2). Ideally, these values should have been derived when the experts defined the uncertainty factors themselves (as the standard deviation can be linked to the experts' level of confidences when defying the uncertainty factors).

This last assumption can be overcome by only using the uncertainty factors coming from the data assessment to derive specific and generic uncertainty factors and so without applying the Bayes theorem. This can be done by simply determining the mean of the uncertainty factors for each subgroup. However, the data harvested for this study don't represent the entire technosphere and this lack of representativeness isn't assessed. Using the Bayes theorem permit to insert a generic information (the experts' judgments); by not using the Bayes theorem, specific information (the one coming from the data assessment) are used to derive generic uncertainty factors that raises the question of representativeness.

Finally, if an access to more data is possible, it will allow the calculation of more refined uncertainty factors, i.e. uncertainty factors more representative of a given sector and for more specific industrial sectors.

6.2.4.2 A way to update not only uncertainty information

Despite the limitations, the developed methodology allowed the calculation of new empirically based uncertainty factors for both the basic and the additional uncertainties that can be directly used or that can be updated as new data become available.

Bayesian inference could also be used in the models used in LCA (whether it is a model to obtain new input parameters for the inventory or models to derive life cycle impact assessment characterisation factors). The inference can especially be used when the parameters are based on temporal series. Updating parameters using inference rather than starting over again the derivation of new parameters allows developers or practitioners to save both computing space (only the posterior information needs to be kept rather than all initial data) and time once the Bayesian model is defined and ready to use that, even if Bayesian Monte Carlo or Markov chain Monte Carlo need to be performed to apply the theorem. Depending on the models, a 5000 steps Monte Carlo simulation takes only few seconds.

CHAPITRE 7 RÉSULTATS COMPLÉMENTAIRES

7.1 Questions soulevées par l'obtention de nouveaux facteurs d'incertitude

Par le développement des nouveaux facteurs d'incertitude présentés au chapitre précédent, des questions se posent notamment sur leurs effets sur les analyses d'incertitude menées.

- Comment varient les distributions résultantes par remplacement des facteurs d'incertitude par ceux développés dans le cadre de ce travail?
- Est-ce que le remplacement des facteurs d'incertitude a un effet sur les conclusions d'une ACV?

Dans la section suivante, une réponse partielle est apportée au premier point ainsi que la méthodologie mise en œuvre pour y répondre.

7.2 Comparaison des analyses d'incertitude menées avant et à la suite de l'obtention de nouveaux facteurs

Afin de déterminer les effets du remplacement des facteurs d'incertitude par ceux développés dans le cadre de ce travail, le remplacement des facteurs d'incertitude additionnelle utilisés jusqu'à présent dans la base de données ecoinvent par les facteurs d'incertitude additionnelle génériques (cf. Tableau 6-3) est réalisé. Une fois ce remplacement effectué, une analyse d'incertitudes pour la catégorie d'impact changements climatiques est menée. Cette analyse d'incertitude consiste en une analyse de Monte Carlo de 5000 pas sur l'ensemble des systèmes de produits modélisés dans la base de données ecoinvent v2.2. Les résultats obtenus sont comparés à ceux obtenus initialement, i.e. sans aucun changement sur les facteurs d'incertitude.

Pour effectuer la comparaison, les moyennes et les écarts-type de chacune des distributions résultantes sont d'abord calculés puis comparés par système de produits. Pour ce faire, le ratio entre la moyenne (respectivement l'écart-type) de la distribution résultante obtenue par

changement de facteurs d'incertitude et la moyenne (respectivement l'écart-type) de la distribution résultante obtenue en conservant les facteurs d'incertitude initiaux est déterminé pour chacun des systèmes de produits. Les distributions de ces ratios sont illustrées à la Figure 7-1 pour la moyenne et à la Figure 7-2 pour l'écart-type. Dans les deux cas, on observe une moyenne et un écart-type de la distribution résultante plus élevés quand les facteurs d'incertitude ont été remplacés. Ce résultat est prévisible puisque, dans la majorité des cas, les facteurs d'incertitude développés dans le cadre de ce travail sont plus grands que ceux appliqués initialement dans la base de données. Ce qui est plus intéressant ici est d'analyser à quel point ces écarts entre respectivement la moyenne et l'écart-type sont importants.

Ainsi, pour les moyennes des distributions résultantes, le ratio moyen est de 1.04. En d'autres termes, l'effet de la mise en application des nouveaux facteurs d'incertitude est donc moindre sur la moyenne de la distribution résultante. Pour ce qui est de l'écart-type, la conclusion n'est pas la même puisque le ratio moyen atteint 1.42. La variabilité de la distribution résultante est donc plus importante en général. Ces faits sont illustrés à la Figure 7-2 où les distributions résultantes pour deux systèmes de produits sont représentées. Cette conclusion ne veut néanmoins pas dire que de l'incertitude a été ajoutée aux résultats d'ACV; mais plutôt que celle-ci a été quantifiée plus « justement » en se basant, certes sur le jugement d'experts mais également sur des données empiriques.

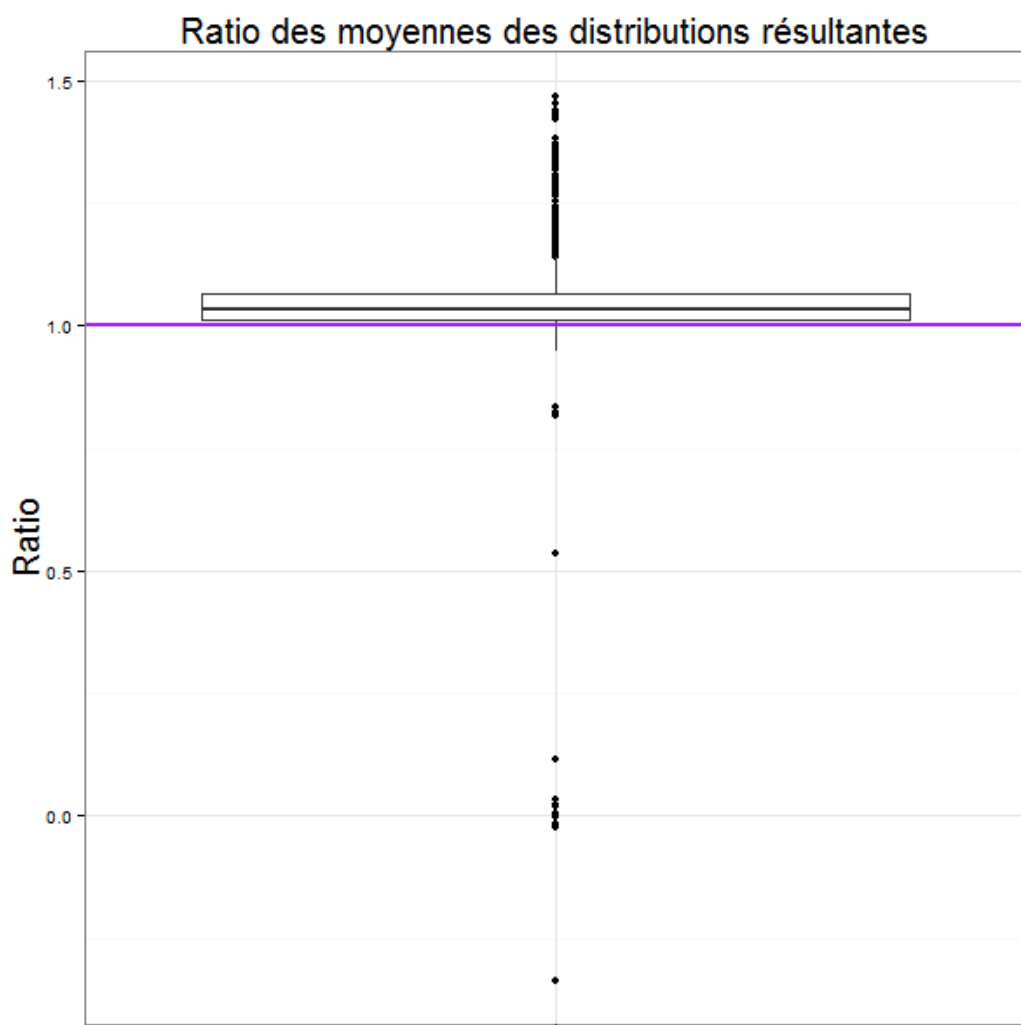


Figure 7-1 Représentation de la distribution du ratio des moyennes des distributions résultantes. Le ratio est effectué entre la moyenne de la distribution résultante obtenue pour le changement des facteurs d'incertitude et la moyenne de la distribution résultante initiale.

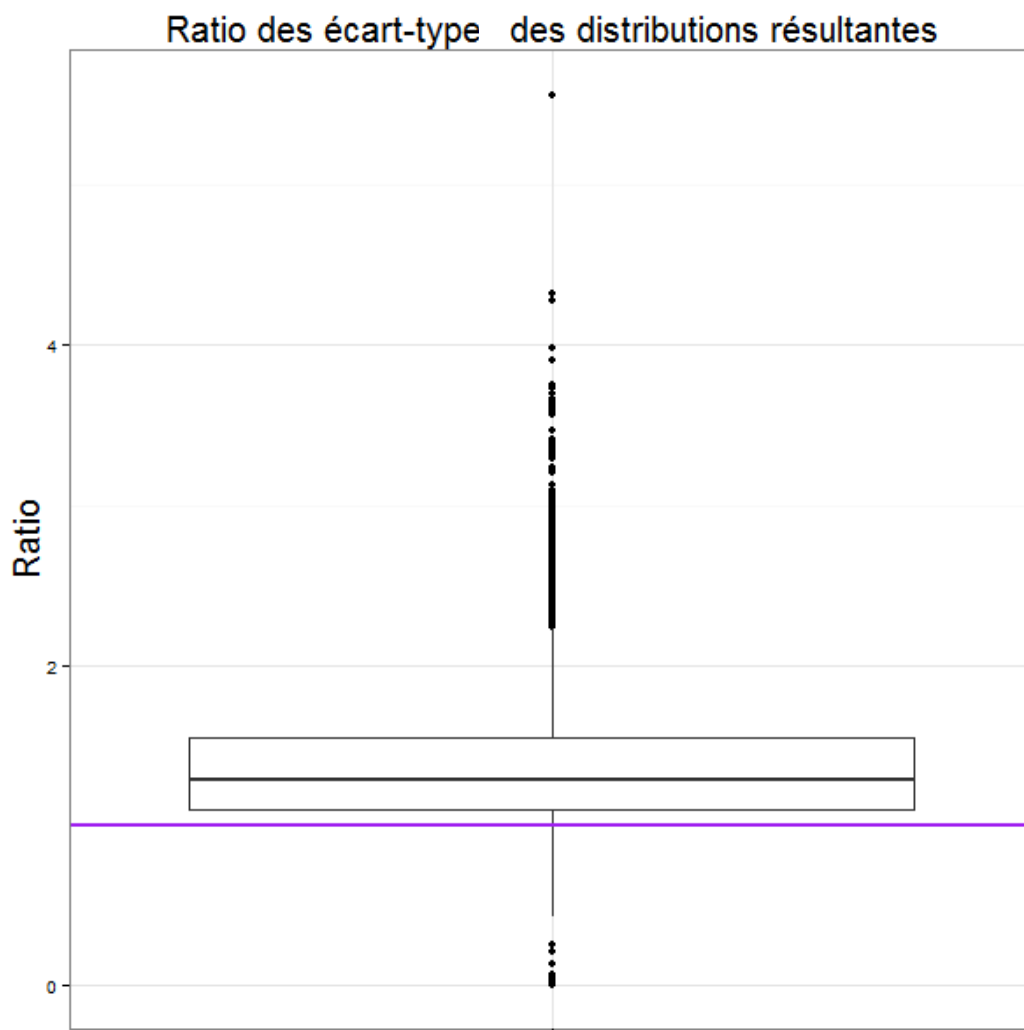


Figure 7-2 Représentation de la distribution du ratio des écarts-type des distributions résultantes. Le ratio est effectué entre l'écart-type de la distribution résultante obtenue pour le changement des facteurs d'incertitude et l'écart-type de la distribution résultante initiale.

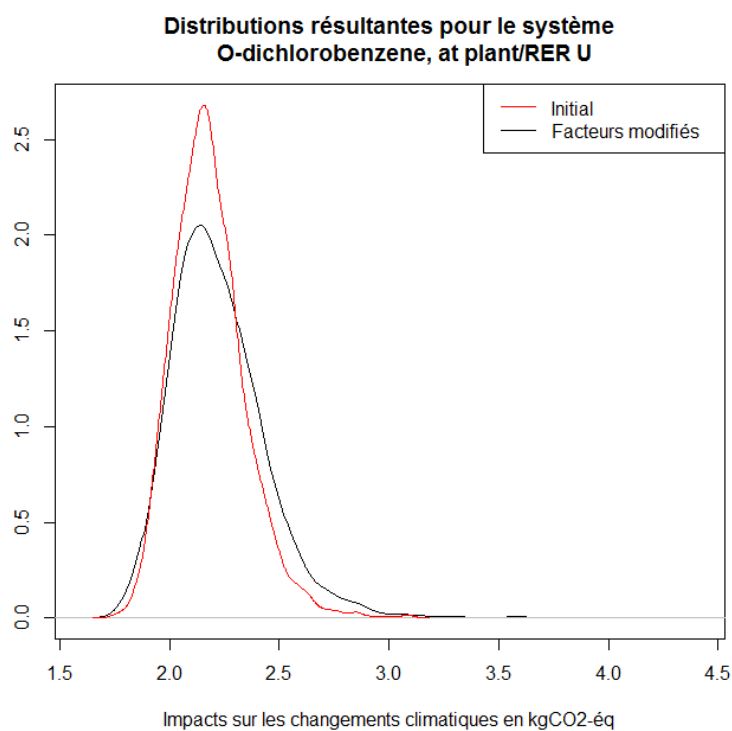
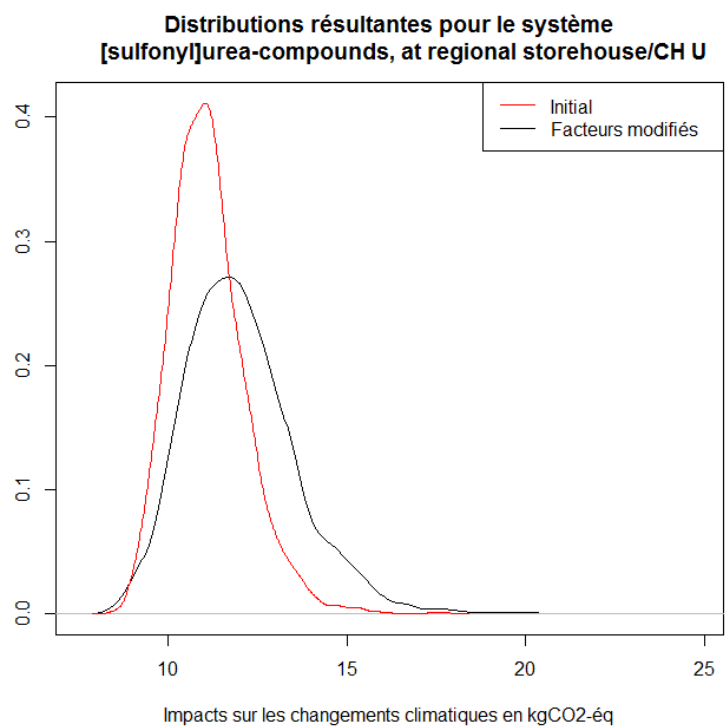


Figure 7-3 Représentation des distributions résultantes de l'impact sur les changements climatiques pour deux systèmes de produits de la base de données ecoinvent v2.2.

CHAPITRE 8 DISCUSSION GÉNÉRALE ET RECOMMANDATIONS

8.1 Atteinte des objectifs de la recherche

De manière globale, l'objectif général de cette thèse qui était de *développer et légitimer par une assise scientifique l'approche pedigree pour son application dans les bases de données d'inventaire* a été atteint par les réponses suivantes aux trois objectifs spécifiques.

1. Légitimer l'usage de la distribution log-normale comme distribution par défaut pour modéliser une donnée d'inventaire avec son incertitude

Les résultats obtenus au Chapitre 4 permettent de répondre à cet objectif. Il a été montré que l'emploi d'une distribution par défaut spécifique pour modéliser une donnée d'inventaire avec son incertitude a une incidence sur les distributions résultantes de l'analyse d'ÉICV. Cette incidence touche en particulier l'asymétrie de la distribution résultante. En ce qui concerne l'effet de la distribution par défaut utilisée sur les résultats liés à la comparaison de deux systèmes, celui-ci est négligeable.

Compte tenu des conclusions obtenues ainsi que des caractéristiques de la distribution log-normale (cf. Tableau 2-3 et Tableau 4-1), la distribution log-normale peut donc être utilisée par défaut pour modéliser une donnée et son incertitude. Néanmoins, si le développeur de processus ou le praticien ont accès à des informations permettant de modéliser une donnée avec son incertitude, ces informations doivent être utilisées en priorité pour définir la distribution de probabilités de cette donnée.

2. Déterminer comment l'approche pedigree telle qu'appliquée actuellement dans la base de données ecoinvent peut être appliquée à des distributions de probabilités autres que log-normale

Les formules présentées au Chapitre 5 permettent de répondre directement à cet objectif spécifique. En appliquant les hypothèses sous-jacentes à l'utilisation de la distribution log-normale pour combiner incertitude de base et incertitude additionnelle, il a été possible de déployer une démarche de développement puis les formules résultantes de cette démarche pour combiner les deux types d'incertitudes pour l'ensemble des distributions présentes dans la base de données ecoinvent pouvant être utilisées pour modéliser une donnée et son incertitude. La démarche développée permettrait de définir des formules de compilation pour n'importe quel type de distribution continue.

Finalement, même si – comme mentionné au paragraphe précédent – la distribution log-normale peut être utilisée par défaut pour modéliser une donnée et son incertitude, il reste pertinent de pouvoir combiner incertitude de base et incertitude additionnelle pour d'autres types de distributions. En effet, si le développeur de données d'inventaire a accès à un échantillon suffisant pour définir la distribution de probabilités propre à la donnée qu'il est en train de modéliser et que cet échantillon ne remplit pas les exigences de qualités, l'ajout de l'incertitude additionnelle permettra de rajouter la variabilité associée à cette qualité imparfaite.

3. Développer une méthode d'analyse de données permettant de déterminer, avec une assise empirique, et de mettre à jour les facteurs d'incertitudes, et ce, par secteur d'activités

Le développement méthodologique présenté au Chapitre 6 permet de définir de nouveaux facteurs d'incertitude empiriques qui se basent à la fois sur le jugement d'experts et sur l'accès à des données spécifiques. Il permet également de mettre à jour ces facteurs si de nouvelles données sont disponibles. Ce troisième objectif spécifique est donc atteint.

De plus, l'application du développement méthodologique a permis d'obtenir des facteurs d'incertitude additionnelle mis à jour, et ce, par secteurs industriels ainsi que des facteurs d'incertitude de base également mis à jour. Ces facteurs sont, pour la majorité, plus élevés que ceux définis auparavant par le jugement d'experts.

Ainsi, les réponses apportées à ces trois objectifs spécifiques ont permis de :

- Justifier l'outil de gestion semi-quantitative de l'incertitude sur les données d'ICV qu'est l'approche pedigree;
- Développer des facteurs d'incertitude permettant de modéliser une donnée d'ICV avec son incertitude de manière plus juste. Ces facteurs d'incertitude peuvent notamment être appliqués aux données d'arrière-plan utilisées dans les bases de données;
- Proposer une approche méthodologique permettant la mise à jour des facteurs d'incertitude.

Si les trois objectifs spécifiques ont été remplis, les développements méthodologiques mis en œuvre pour répondre à ces derniers reposent sur un certain nombre d'hypothèses. Ces hypothèses et leurs effets possibles sur les résultats font en sorte que les conclusions qui ont été tirées précédemment doivent être interprétées avec humilité tel que discuté à la section qui suit.

8.2 Les limites des résultats obtenus

8.2.1 Des hypothèses sur la modélisation des facteurs d'incertitude

L'ajout de l'incertitude de base à l'incertitude additionnelle ainsi que la détermination des facteurs d'incertitude passe par la modélisation de ces derniers; modélisation qui implique quelques choix. Les effets de ces choix ne sont pas quantifiés ici, ils sont plutôt caractérisés qualitativement expliquant d'où l'incertitude de modélisation des facteurs d'incertitude pourrait provenir.

Un des premiers choix effectués est celui du paramètre représentant le facteur d'incertitude. Il a été mentionné précédemment que différents paramètres avaient été proposés par les auteurs, parmi eux le CV, le carré de l'écart-type géométrique et l'écart-type logarithmique (cf. section 3.3.2). Dans le cadre de cette thèse, le CV a été utilisé pour l'application de l'approche pedigree à l'ensemble des distributions de probabilités (cf. Chapitre 5) et le GSD² pour le développement de facteurs d'incertitude réalisé dans le cadre du troisième article (cf. Chapitre 6). Bien qu'il n'ait pas d'incidence sur les résultats en eux-mêmes, ce manque de cohérence autour de ces deux choix peut avoir une incidence sur la clarté des résultats présentés. En appliquant le lien existant

entre le CV et le GSD^2 pour une distribution log-normale (cf. Tableau 5-0-2), les facteurs d'incertitude peuvent facilement être exprimés en termes de CVs et donc être directement appliqués avec les résultats obtenus au deuxième article.

Comme stipulé à la phrase précédente, il est supposé ici que les facteurs d'incertitude suivent une distribution log-normale. Le choix de cette distribution joue un rôle important notamment pour l'application du théorème de Bayes qui permet ici l'obtention de facteurs d'incertitude. En effet, le choix de la distribution à priori a une incidence pour l'application du théorème et, suivant son type, peut influencer la distribution à posteriori (Morita, et al., 2010). La distribution log-normale utilisée à priori permet, puisqu'elle est à support de définition semi-infini, de ne discriminer aucune valeur positive possible pour le facteur d'incertitude. Par des applications successives du théorème de Bayes, l'influence initiale de l'information à priori tend de plus à diminuer.

Afin de tester l'influence de cette distribution à priori, le théorème de Bayes est appliqué une seconde fois pour certains facteurs d'incertitude additionnelle génériques avec le choix d'une distribution à priori uniforme. Comme les distributions uniforme et log-normale (représentant la fonction de vraisemblance) ne sont pas conjuguées, le théorème de Bayes doit être appliqué de manière numérique. Pour ce faire, les techniques de Monte-Carlo par chaînes de Markov (MCMC) sont utilisées en utilisant le logiciel OpenBUGS. Le minimum et le maximum de la loi uniforme sont déterminés en fonction des données de l'Annexe D. Les résultats obtenus pour une simulation de 5000 pas sont comparés avec ceux obtenus au troisième article (cf. Chapitre 6) et sont présentés au Tableau 8-1. Ces résultats montrent que le type de distribution utilisé à priori a une incidence sur les résultats. Cette incidence est due à la forme de la distribution uniforme, le fait qu'elle soit bornée fait en sorte que la distribution à posteriori ne peut être définie à l'extérieur de ces bornes. Ceci est illustré à la Figure 8-1. Compte tenu du fait que la manière dont le jugement d'experts a été élicité pour le développement de facteurs d'incertitude initiaux n'est pas documentée et qu'aucune information sur leurs valeurs minimale et maximale n'est donnée, le choix d'une distribution non bornée est justifiée ici. Néanmoins, avec ce choix vient une hypothèse liée à l'incertitude sur le facteur d'incertitude et plus spécifiquement sur la valeur de l'écart-type logarithmique servant à modéliser la distribution log-normale à priori.

Tableau 8-1 Résultats de l'analyse bayésienne par MCMC pour une fonction à priori uniforme et comparaison avec les résultats obtenus à l'article 3. La *fonction à posteriori1* représente la fonction obtenue à l'article 3 et la *fonction à posteriori2* représente celle obtenue dans le cas d'une loi à priori uniforme.

Indicateur pedigree	Score	Paramètre de la loi uniforme utilisée à priori		Facteur d'incertitude générique issue de la fonction à posteriori 1	Facteur d'incertitude générique issue de la fonction à posteriori 2	Erreur relative
		min	max			
Justesse	2	1.05	1.54	1.01	1.05	3.96%
	3	1.02	1.64	1.21	1.37	13.22%
	4	1.20	1.69	1.25	1.42	13.60%
	5	1.50	5.10	2.36	3.81	61.44%
Corrélation temporelle	2	1.00	1.22	1.09	1.11	1.83%
	3	1.00	1.73	1.55	1.58	1.94%
	4	1.01	2.88	2.22	2.49	12.16%
	5	1.01	3.86	2.49	3.22	29.32%
Corrélation géographique	2	1.00	1.40	1.14	1.26	10.53%
	3	1.02	1.64	1.23	1.41	14.63%
	4	1.03	2.57	2.36	2.32	1.69%
	5	1.10	2.57	2.30	2.32	0.87%
Corrélation technologique	2	1.00	1.37	1.19	1.26	5.88%
	3	1.02	1.65	1.52	1.50	1.32%
	4	1.07	2.08	1.95	1.88	3.59%
	5	2.00	2.80	2.23	2.54	13.90%

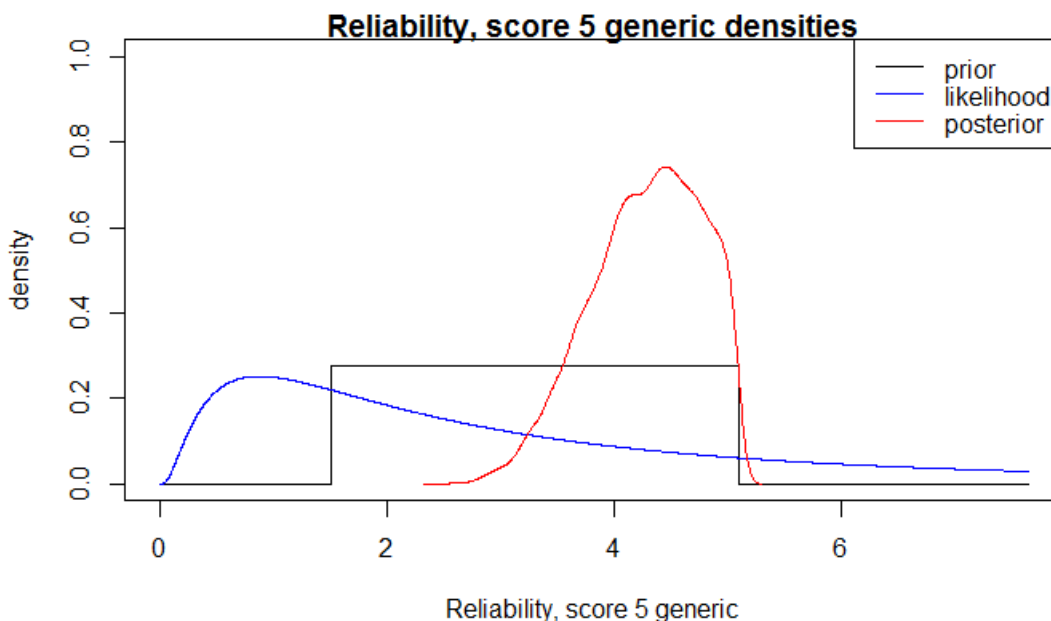


Figure 8-1 Représentation des distributions à priori (*prior*) uniforme et à posteriori (*posterior*) ainsi que de la fonction de vraisemblance (*likelihood*) pour le score 5 de l'indicateur justesse

8.2.2 La représentativité des données utilisées pour développer de nouveaux facteurs d'incertitude

Les données utilisées pour développer les facteurs d'incertitude proposés à l'article 3 sont issues de nombreuses sources de données. Si au moment de la collecte de ces données, un effort particulier a été fait pour accéder à des données représentatives de l'ensemble des secteurs industriels, cette représentativité peut être discutée. La Figure 8-2 témoigne de ce manque de représentativité. Pour chaque secteur industriel spécifique, le pourcentage de sous-secteurs déterminés par la classification NAICS au niveau 6 est représenté, au mieux c'est près de 60% de ces sous-secteurs qui sont représentés dans l'analyse menée. Ce manque de représentativité est lié à l'accessibilité de ces données. Dans ce projet, les données ont été collectées à la main (ce qui peut aussi entraîner des erreurs dans leur traitement) au travers des sources à disposition. L'accès et la collecte de données d'inventaire –libres- reste encore aujourd'hui une problématique en ACV (Steinmann et al., 2014). En plus d'une limite dans la représentativité, ce difficile accès aux données a fait en sorte que l'approche, et notamment le développement des facteurs d'incertitude,

n'a pu être réalisée pour certains types de données utilisées en ICV (par exemple les paramètres permettant de relier une propriété physique, économique ou temporelle à des flux d'inventaire).

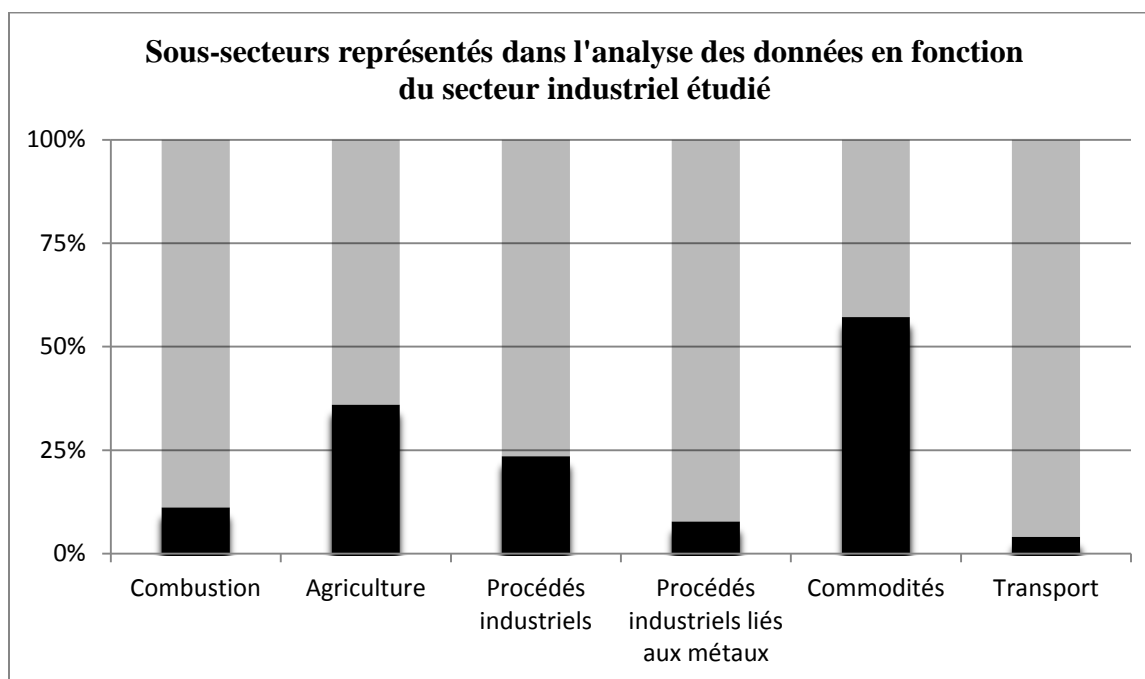


Figure 8-2 : Pourcentages des sous-secteurs NAICS définis au niveau 6 représentés dans l'analyse de données pour chaque secteur industriel spécifique

L'utilisation de la méthode bayésienne permet néanmoins de s'affranchir partiellement de cette limite puisque l'accès à de nouvelles ou à de « meilleures » données permettra de mettre à jour les facteurs d'incertitude obtenus et de les raffiner. Les résultats présentés ici ne doivent donc pas être considérés comme définitifs mais plutôt comme les premiers pas d'un processus itératif qui permet le développement de facteurs d'incertitude qui représenteront de mieux en mieux la réalité qu'ils modélisent.

8.2.3 Des facteurs d'incertitude additionnelle surestimés?

Lors de la détermination des facteurs d'incertitude additionnelle, le nombre de données collectées n'a pas été suffisant pour les déterminer de manière totalement parfaite (mis à part le critère de complétude pour lequel cette remarque ne s'applique pas). En effet, pour déterminer l'incertitude

additionnelle due par exemple exclusivement à la corrélation temporelle, des données ne possédant aucun autre type de variabilité (qu'elle soit technologique, géographique ou liée à la manière dont les données ont été obtenues) ne devrait être présente. Si les sous-groupes pour la détermination de l'incertitude additionnelle pour la corrélation temporelle ont été créés en tenant compte du secteur industriel, la zone géographique et la manière dont les données ont été obtenues n'ont, quant à eux, pas été fixés. Les facteurs d'incertitude additionnelle pour le critère de corrélation temporelle possèdent donc une part de variabilité due aux critères de justesse et de corrélation géographique. Néanmoins, dans les faits, souvent les sous-groupes créés sont composés de données issues d'une même aire géographique et basées sur les mêmes moyens d'obtention.

Les Figure 8-3, Figure 8-4 et Figure 8-5 illustrent ce fait. Elles présentent les valeurs du GSD^2 calculées pour les 5 scores de corrélation technologique et en considérant les différents scores que ces données prendraient pour les autres indicateurs (respectivement les indicateurs de justesse – Figure 8-3, de corrélation géographique – Figure 8-4 et de corrélation temporelle – Figure 8-5). Dans un cas parfait, les GSD^2 pour le critère de corrélation technologique ne dépendraient pas d'un autre indicateur, or ce n'est pas le cas ici à part pour le score 4 de corrélation technologique comparé au critère de justesse. Néanmoins, cette variabilité résiduelle dans les GSD^2 est diminuée en déterminant les facteurs d'incertitude (cf. Equation 6-7).

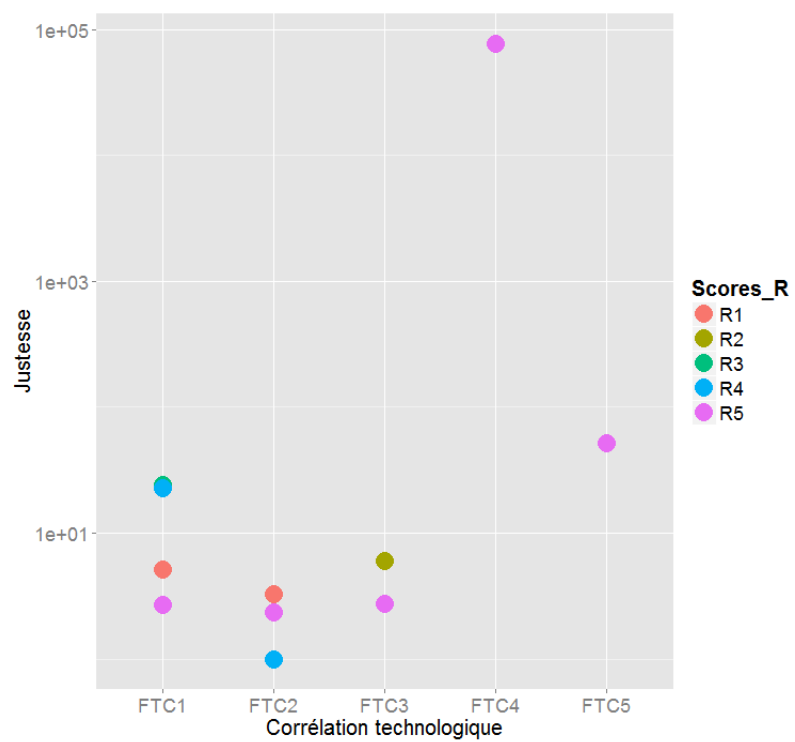


Figure 8-3 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de justesse (R) et de corrélation technologique (FTC) sont étudiés simultanément

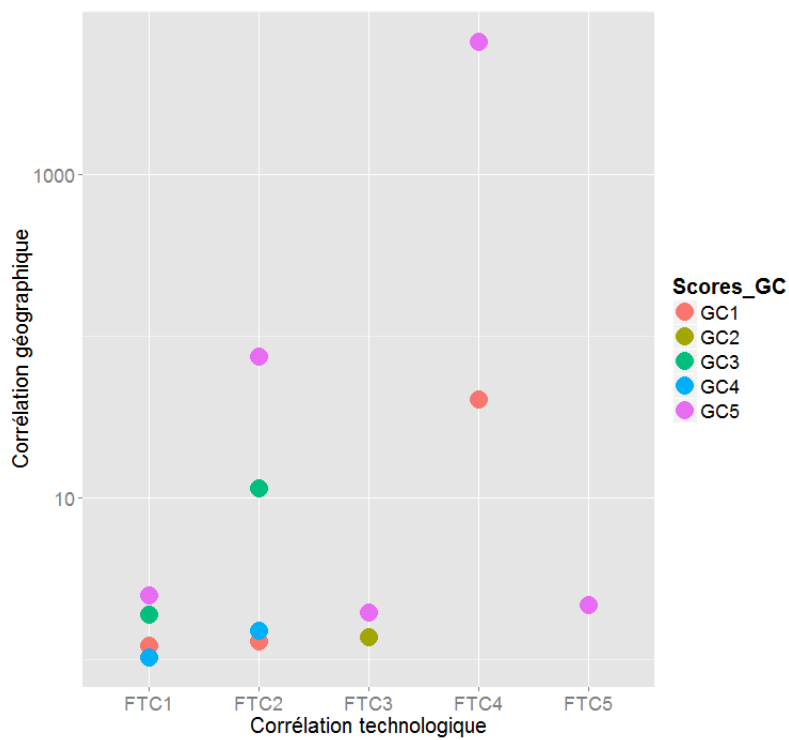


Figure 8-4 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de corrélation géographique (GC) et de corrélation technologique (FTC) sont étudiés simultanément

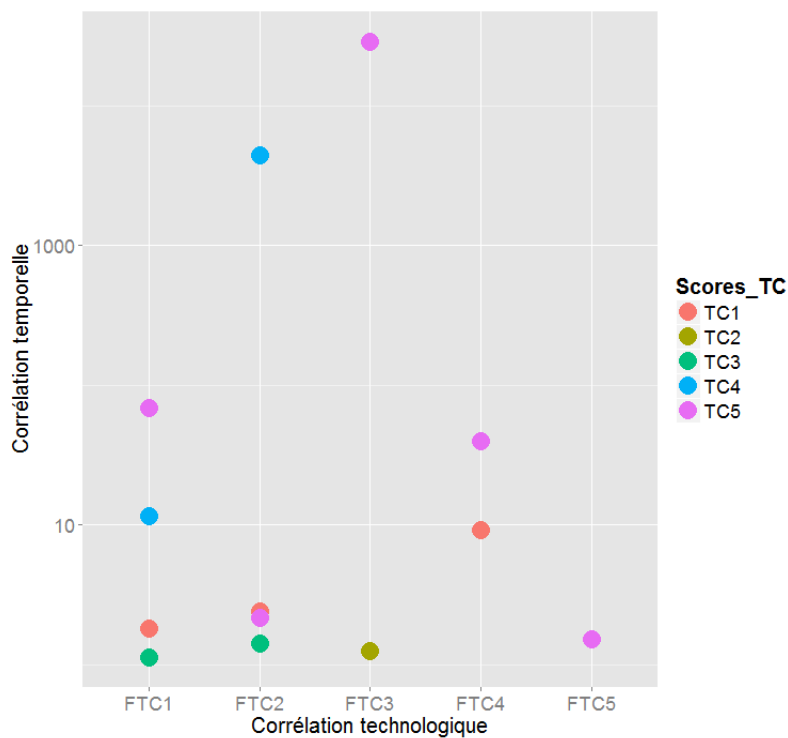


Figure 8-5 Représentation du carré de l'écart-type géométrique calculé pour chaque couple indicateur/score pedigree quand les indicateurs de corrélation temporelle (TC) et de corrélation technologique (FTC) sont étudiés simultanément

8.3 Recommandations

8.3.1 Recommandations pour de futures recherches

8.3.1.1 S'affranchir des limites précédemment citées et poursuivre l'amélioration de l'approche pedigree

Tel que mentionné précédemment, les travaux réalisés s'inscrivent dans un processus itératif qu'il est possible de raffiner en traitant des points suivants.

1. Déterminer des facteurs d'incertitude additionnelle ne possédant pas de variabilité résiduelle

Afin de déterminer des sous-groupes de données permettant de ne considérer qu'un unique critère pedigree, un accès à des données à la fois plus spécifiques et représentatives peut-être envisagé. Si les données ne sont pas disponibles ou accessibles, il est possible d'étudier les effets des variables qualitatives que sont les indicateurs pedigree sur la variable quantitative qu'est le facteur d'incertitude. Cette étude permettrait de quantifier la part de la variabilité des facteurs d'incertitude pouvant être expliquée par chaque indicateur pedigree ou par les interactions entre indicateurs et ainsi de distribuer ces effets d'interactions pour ne pas potentiellement surestimer la valeur du facteur d'incertitude total.

2. Une automatisation possible de la collecte de données?

Une des limites citées à l'accès aux données est la collecte à la main de ces dernières. Cette collecte manuelle demande un temps important et n'a pas permis ici une collecte systématique de l'information à disposition. La question suivante se pose donc : y a-t-il une manière d'automatiser cette collecte ou du moins de la rendre plus systématique? Ceci pourrait passer par exemple par des techniques d'extraction de l'information adaptées au cas particulier des typologies de données traitées dans ce projet. Les techniques d'extraction pourraient avoir ici plusieurs fonctions afin d'aider à la collecte et à l'analyse des données, fonctions présentes dans le cadre d'applications dans d'autres domaines (Grivel, 2001) :

- Automatiser la construction de la base de données;
- Spécifier le type de données présentes dans chaque source de données afin de faciliter le traitement par la suite.

Un autre point touchant la collecte de données, notamment pour la mise à jour des facteurs d'incertitude, et qui pourrait être mis en place avec les techniques d'extraction de la connaissance serait la mise en place d'une liste de sources de données modifiées annuellement (ou sur une base temporelle plus grande). Parmi ces sources de données on trouve notamment les bases de données des facteurs d'émission ou des bases de données plus spécifiques à un secteur particulier.

3. La modélisation de la distribution des facteurs d'incertitude

La forme de la distribution à priori du facteur d'incertitude a un effet sur la détermination de ces facteurs d'incertitude après application du théorème de Bayes. Afin d'avoir une base solide pour la distribution à priori, il serait possible de valider les facteurs d'incertitude développés ici en appliquant des techniques d'élicitation du jugement d'experts pour ces derniers. L'élicitation du jugement d'experts consiste à regrouper différents experts afin de les interroger sur la forme et/ou les valeurs spécifiques qui devraient être prises par la distribution à priori. À partir de ces informations, le statisticien va pouvoir définir la distribution adéquate (O'Hagan, 2006). En plus d'utiliser leur jugement pour la définition de la distribution à priori, les experts pourraient également donner leur avis sur les facteurs d'incertitude développés dans le cadre de ce travail et ainsi valider ces derniers en se basant sur leur domaine d'expertise (domaine propre à un secteur industriel spécifique par exemple).

Afin de développer ces trois points, une seule expertise en ACV ne suffit pas, il s'agit de regrouper des expertises touchant à différents domaines : de l'analyse de données multi-variées à la psychologie et l'animation de panels inclusifs en passant par de l'analyse textuelle. De plus, ces points ne touchent que l'approche développée ici (même si d'autres applications dans le domaine de l'ACV pourraient leur être trouvées); d'autres problématiques reliant incertitudes et ACV pourraient également être mentionnées.

4. La modification de la valeur déterministe

Le premier article faisant mention de l'approche pedigree appliquée à l'ACV propose également de modifier la valeur déterministe par ajout de l'incertitude additionnelle, et dans certains cas, des pourcentages d'augmentation de cette valeur sont donnés (Weidema & Wesnæs, 1996). L'analyse de données effectuée ici pourrait également permettre de mettre à jour les valeurs déterministes, que ce soit par une augmentation ou une diminution de cette dernière. En effet, des séries temporelles ont par exemple été analysées pour la détermination de facteur d'incertitude additionnelle, mais elles auraient également pu être utilisées en analysant la variation des valeurs d'échanges à travers le temps. Ont-elles diminué par l'augmentation de l'efficacité des processus ou par durcissement de la législation? Si oui, les données collectées ainsi que l'utilisation de l'inférence bayésienne pourraient être utilisées pour la mise à jour.

8.3.1.2 Incertitudes et ACV

Actuellement, l'inférence bayésienne est très peu utilisée en ACV (cf. section 2.3.2). Pourtant, de par ses caractéristiques, son utilisation permettrait d'améliorer les modèles et/ou les données entrées de ces modèles utilisés en ACV et ainsi obtenir des résultats plus représentatifs. L'inférence pourrait être utilisée pour les données d'ICV, pour les modèles développés pour l'ICV ou encore, par exemple, pour mettre à jour les facteurs de caractérisation utilisés en ÉICV.

La prise en compte de l'incertitude en ACV pour la prise de décision ne semble pas encore être réalisée de manière systématique. Dans le même ordre d'idée, très peu d'études ACV publiées font mention d'une analyse d'incertitudes et présentent ces dernières. Cette faible considération de l'incertitude peut également être visible en recherche où souvent l'incertitude n'est considérée que comme une plus-value d'un développement méthodologique plutôt que d'être considérée parallèlement au développement.

Pourquoi, malgré les outils existant actuellement, l'incertitude est-elle encore si peu considérée en ACV? Dans l'ouvrage « Penser l'incertain », les auteurs lient le concept d'incertitude à celui de risque et avancent qu'une fois que la notion de risque est communiquée un sentiment de non maîtrise de l'incertain peut survenir (Vrancken, 2014). Même si en ACV, la notion de « gestion du risque » est très peu présente (elle est plus traitée comme une externalité et gérée par les preneurs de décisions), est-ce que ce lien pourrait expliquer le peu de considération de l'incertitude? Lors d'une étude réalisée pour le compte de l'institut national de la santé publique et de l'environnement néerlandais sur la prise en compte des incertitudes dans un outil de quantification des risques liés à l'utilisation de produits chimiques, beaucoup d'utilisateurs de l'outil sont préoccupés par le fait que si un résultat est présenté comme incertain, le processus de décision en sera ralenti et aucune décision ne sera prise (Jager, 1998). Est-ce que cette préoccupation existe aussi chez les praticiens de l'ACV? Il s'agirait de comprendre les causes de la faible considération de l'incertitude afin d'augmenter cette dernière puisque les avantages en seraient nombreux, parmi eux :

- Augmenter la confiance à accorder à un résultat;
- Justifier le fait de devoir ou non compléter une analyse afin de pouvoir conclure sur les résultats (Jager, 1998);

- Améliorer l'interprétation et la connaissance en apprenant des informations que la gestion de l'incertitude nous livre (Vrancken, 2014).

En déterminant ces causes, en inventant de meilleurs moyens de communiquer l'incertitude et en sensibilisant davantage la communauté scientifique, les preneurs de décisions et les utilisateurs en général à la gestion de l'incertitude, la prise en compte de cette dernière pourrait être démocratisée. Ce travail implique également la collaboration entre différentes disciplines (des sciences sociales aux professionnels de la communication) et pourrait permettre de s'affranchir de l'inconfort relié à la communication et à la gestion de l'incertitude en ACV et en sciences de manière générale.

8.3.2 Comment intégrer les travaux dans les bases de données?

Une partie des résultats obtenus dans le cadre de ce projet peuvent être appliqués dans les bases de données utilisant la structure de gestion de l'incertitude qu'est l'approche pedigree (soit la base de données ecoinvent) ou encore dans les logiciels supportant de telles bases de données. Ces résultats sont :

- Les formules d'application de l'approche pedigree aux autres distributions de probabilités (cf. Chapitre 5);
- Les différents facteurs d'incertitude développés pour l'incertitude de base et l'incertitude additionnelle (cf. Chapitre 6).

Pour ce qui est de l'implémentation du premier point, cela devra passer par la programmation des différentes formules à la place de l'unique formule de compilation de l'incertitude de base et incertitude additionnelle qui est disponible actuellement.

En ce qui concerne l'intégration des facteurs d'incertitude développés dans le cadre de ce projet, celle-ci dépend du type de facteurs. Pour les facteurs d'incertitude de base et les facteurs d'incertitude additionnelle génériques, il suffit de remplacer les valeurs jusque-là utilisées par ceux ici développés. Une préparation préliminaire est toutefois à réaliser en ce qui concerne les facteurs d'incertitude additionnelle développés pour des secteurs industriels spécifiques. Les processus présents dans la base de données ecoinvent sont identifiés par secteur industriel en

utilisant la classification ISIC (classification internationale des secteurs industriels) (Ecoinvent, 2014), la classification utilisée dans le cadre de ce projet est la classification nord-américaine NAICS. Il s'agit donc dans un premier temps de lier les classes de la classification ISIC à celles de la NAICS. Ces liens ont d'ores et déjà été définis notamment par la division statistiques des Nations Unies (UN Statistics Division, 2011). Une fois ces liens réalisés, il s'agit :

- Pour les flux élémentaires : d'associer les facteurs d'incertitude spécifiques au secteur industriel spécifique auquel le processus associé appartient;
- Pour les flux économiques : d'associer les facteurs d'incertitude spécifiques à leur classe NAICS spécifique.

Idéalement, et si la demande est présente, les développeurs de bases de données ou d'outils permettant de réaliser des ACVs en se basant sur ces bases de données pourraient également intégrer des outils plus avancés permettant par exemple d'intégrer l'inférence à l'étude qui est menée, de proposer de meilleures interfaces pour comprendre et intégrer l'incertitude, de permettre une communication de l'incertitude avec de meilleurs outils graphiques.

Le but de ces outils devrait, certes, être de permettre une gestion de l'incertitude mais surtout de permettre de démocratiser sa communication par les praticiens de l'ACV afin qu'elle soit considérée par les preneurs de décisions.

8.3.3 Comment utiliser les travaux ici développés?

Par une intégration des travaux dans les bases de données, que ce soit ecoinvent ou une autre, l'utilisation régulière des travaux par le praticien ou le développeur de données en serait facilitée, car quasi-automatique. En effet, actuellement l'utilisation de l'approche pedigree dans une étude ACV se fait pour le praticien et le développeur en renseignant les champs suivants : « type de distribution », « incertitude de base » et « quintuplet pedigree ». Si la distribution choisie est log-normale, l'incertitude de base et de l'information issue du quintuplet pedigree sont combinées. Ainsi, l'intégration des travaux dans les bases de données, permettrait aux praticiens et aux développeurs de pouvoir choisir la meilleure distribution pour modéliser une donnée et son incertitude de base ainsi que de lui ajouter l'incertitude additionnelle en renseignant la matrice

pedigree. Ce renseignement de la matrice permet également, connaissant le type d'échanges modélisés, de sélectionner le facteur d'incertitude propre au secteur industriel étudiée.

Si les travaux ne sont pas intégrés dans les bases de données ou les logiciels permettant de réaliser des ACVs, l'utilisation des résultats demande une manipulation de ces derniers de la part des praticiens souhaitant les utiliser ainsi que la modélisation du système étudié avec des logiciels d'analyse numérique (de type R, Python, Matlab,...) qui supportent la conduite d'analyse d'incertitude par analyses de Monte Carlo par exemple. La modélisation du système dans ce type de logiciels est nécessaire si le praticien souhaite utiliser l'ensemble des résultats obtenus à la fois sur les données d'avant-plan et d'arrière-plan. Si seule la modélisation de l'incertitude sur les données d'avant-plan importe (ce qui est le cas par exemple pour un développeur), il est néanmoins possible d'intégrer les résultats présentés en modélisant l'incertitude sur une donnée « à la main » avant de l'intégrer au système modélisé. Pour ce faire, il s'agit, pour chaque donnée, de choisir le type de distribution la modélisant avec son incertitude, de combiner incertitude de base et incertitude additionnelle en employant les équations présentées au Chapitre 5 en utilisant les facteurs d'incertitude adéquats puis de rentrer la donnée avec sa distribution et son incertitude totale dans le logiciel utilisé pour que cette dernière soit considérée dans l'analyse d'incertitude.

CHAPITRE 9 CONCLUSION

9.1 Rappel des résultats et de l'originalité du projet

En répondant à l'objectif de recherche de développer et légitimer par une assise scientifique la structure de gestion semi-quantitative de l'incertitude sur les données d'inventaire qu'est l'approche pedigree, ce projet de recherche a permis de s'affranchir des limites qui étaient actuellement citées autour de l'approche pedigree. Ainsi,

- La justification du choix de la distribution log-normale comme distribution par défaut pour modéliser une donnée et son incertitude;
- L'application de l'approche pedigree aux distributions autres que log-normales disponibles dans la base de données ecoinvent pour modéliser une donnée et son incertitude;
- Le développement d'une méthodologie permettant la mise à jour des facteurs d'incertitude;
- Le développement de facteurs d'incertitude empiriques pour l'ensemble des couples « indicateur pedigree – score pedigree » et de certains facteurs d'incertitude de base

ont permis- en plus de s'affranchir des limites - pour la première fois, la justification quasi-complète de l'approche pedigree telle qu'appliquée dans la base de données ecoinvent.

Certes, ce projet en lui-même repose sur certaines limites notamment reliées aux choix réalisés pour mettre en œuvre les développements méthodologiques; néanmoins il permet de répondre à certains écueils. En plus des résultats présentés et des conclusions tirées, les éléments suivants ont pour la première fois été traités au cours de ce travail:

1. La méta-analyse menée sur l'ensemble des systèmes de produits de la base de données ecoinvent permettant de tester l'effet d'un changement de distribution utilisée pour modéliser une donnée et son incertitude;
2. L'approche analytique développée pour l'application de l'approche pedigree aux autres types de distribution;
3. La modélisation de l'incertitude sur l'incertitude et, au-delà, de ce point l'utilisation de l'inférence bayésienne pour modéliser les facteurs d'incertitude.

9.2 Intégration des résultats obtenus et futurs développements

Pour aller plus loin dans la gestion des incertitudes via l'approche pedigree, les futurs développements à envisager devront permettre, dans un premier temps, de s'affranchir des limites citées au chapitre précédent. Ceci passe notamment par l'accès et la collecte d'un nombre plus important de données ainsi que par une meilleure modélisation de la distribution des facteurs d'incertitude. Un accès systématique à de nouvelles données permettra également de perpétuer la mise à jour des facteurs d'incertitude en utilisant la méthodologie développée ici. L'enjeu, pour ce faire, est donc principalement un enjeu d'accès d'accessibilité aux données. Celui-ci pourrait être en partie diminué en utilisant des outils statistiques.

Néanmoins, malgré ses limites, l'approche pedigree a le principal avantage de permettre la prise en compte systématique de l'incertitude sur les données d'inventaire. Si les travaux développés ici l'ont été pour ce cas particulier de l'incertitude sur les données d'ICV, les résultats obtenus et les approches méthodologiques développées peuvent également être appliqués et implémentés à d'autres cas faisant intervenir l'approche pedigree. L'approche a, par exemple, été envisagée pour être intégrée dans les travaux reliés au développement de la méthodologie d'évaluation des impacts IMPACT World+ en cours de développement. L'approche pedigree utilisée dans ce cas aurait pu permettre une évaluation semi-quantitative de l'incertitude sur les facteurs de caractérisation. Les lignes directrices pour la réalisation d'ACVs sociales parues en 2009 font également mention du développement possible de l'approche pedigree pour analyser les incertitudes des données utilisées en ACV sociales (UNEP, 2009).

Les développements réalisés dans le cadre de ces travaux n'ont d'intérêt que si les résultats qui en découlent sont utilisés. Chaque personne en ayant besoin ou voulant les utiliser peut le faire, mais leur intégration directe dans les bases de données ou les logiciels supportant les bases de données permet une utilisation systématique de ces résultats. À ce titre, des échanges ont eu lieu avec les gestionnaires de la base de donnéesecoinvent qui souhaitent intégrer les facteurs d'incertitude obtenus lors de la sortie de la prochaine version de la base de donnéesecoinvent. Des discussions ont également eu lieu avec les développeurs du logiciel libre OpenLCA qui, eux aussi, souhaitent intégrer les facteurs d'incertitude développés au sein du logiciel.

9.3 Démocratiser la prise en compte de l'incertitude et de la variabilité en ACV

Que ce soit de par les résultats ici publiés ou par d'autres travaux menés dans différents groupes de recherche, la connaissance sur l'incertitude et sa prise en compte ainsi que l'expression de la variabilité en ACV s'améliorent. Par exemple, ici, une approche systématique de la prise en compte des incertitudes sur les données d'ICV est justifiée et améliorée. Ces développements peuvent être démocratisés certes, comme on l'a vu précédemment, par l'intégration des travaux dans les bases de données ou les logiciels ACV, mais l'intégration ne veut pas forcément dire une utilisation systématique des résultats de ces travaux.

Un exemple de l'utilisation systématique des résultats obtenus ici, et notamment des facteurs d'incertitudes empiriques élaborés, pourrait par exemple être illustré par une conduite systématique d'une analyse d'incertitude pour chaque ACV menée. Ceci n'est pas fait actuellement pour différentes raisons, notamment le temps de calcul nécessaire ou encore le manque de ressources pour modéliser l'incertitude sur les données d'avant-plan par exemple. La compréhension des enjeux autour de la prise en compte des incertitudes en ACV ainsi qu'autour de la communication de ces dernières semble également être une avenue de recherche pour la démocratisation de l'incertitude en ACV et dans les autres outils d'aide à la décision.

En présentant un résultat, qu'il soit absolu ou relatif, sans avoir tenu compte des incertitudes et de la variabilité, qui sont une connaissance en soi sur les données et les modèles employés, n'y-a-t-il pas une partie de l'information connue sur les résultats omise pour le décideur? En améliorant la modélisation de l'incertitude, en définissant de manière plus précise les sources de variabilité et leurs effets, en développement de meilleurs moyens de communiquer l'incertitude et de sensibiliser le public à sa prise en compte, peut-être que la considération de l'incertitude pourrait être réalisée de manière plus systématique et être acceptée non pas comme une faiblesse des résultats présentés, mais comme une source de connaissance en soi sur ces résultats.

BIBLIOGRAPHIE

- André, J., & Lopes, D. (2012). On the use of possibility theory in uncertainty analysis of life cycle inventory. *The International Journal of Life Cycle Assessment*, 17(3), 350-361. doi:10.1007/s11367-011-0364-9
- Basson, L., & Petrie, J. G. (2007). An integrated approach for the consideration of uncertainty in decision making supported by Life Cycle Assessment. *Environmental Modelling & Software*, 22(2), 167-176. doi:http://dx.doi.org/10.1016/j.envsoft.2005.07.026
- Bebbington, M. Lai, C.D.; & Zitakis, R. (2008). Lifetime analysis of incandescent lamps: the Menon-Agrawal model revisited. *Reliability & Risk Analysis - Theory and Applications*, 1(1), 97-108. Retrieved from http://gnedenko-forum.org/Journal/2008/012008/RATA_1_2008-10.pdf
- Ben Letham, C. R. (2012). 15.097 Probabilistic Modeling and Bayesian analysis 15.097 Prediction: Machine Learning and Statistics - Spring 2012: MIT OpenCourseWare.
- Björklund, A. (2002). Survey of approaches to improve reliability in LCA. *The International Journal of Life Cycle Assessment*, 7(2), 64-72. Retrieved from <http://dx.doi.org/10.1007/BF02978849>
- Blonk H., Ponsioen T, Kool, A, Marinussen, M. (2011). *The Agri-footprint method - Methodological LCA framework, assumptions and applied data - Annex*. Retrieved from The Netherlands:
- Brundtland, G. H., & CMED. (1987). *Notre avenir à tous*: Commission Mondiale sur l'Environnement et le Développement (CMED).
- Budescu, D., Por, H.-H., & Broomell, S. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change*, 113(2), 181-200. doi:10.1007/s10584-011-0330-3
- Bulle C.; Humbert S. ; Margni,M. ; Rosenbaum,R. ; Joliet,O. (2012). IMPACT World+ official launch *Presentation during the 6th SETAC World Congress*. Berlin.
- Canter, K., Kennedy, D., Montgomery, D., Keats, J. B., & Carlyle, W. M. (2002). Screening stochastic Life Cycle assessment inventory models. *The International Journal of Life Cycle Assessment*, 7(1), 18-26. doi:10.1007/bf02978906
- Ciroth, A. (2009). Cost data quality considerations for eco-efficiency measures. *Ecological Economics*, 68(6), 1583-1590. doi:DOI: 10.1016/j.ecolecon.2008.08.005
- Ciroth, A., Fleischer, G., & Steinbach, J. (2004). Uncertainty calculation in life cycle assessments. *The International Journal of Life Cycle Assessment*, 9(4), 216-226. Retrieved from <http://dx.doi.org/10.1007/BF02978597>
- Ciroth, A., Muller, S., Weidema, B., & Lesage, P. (2013). Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *The International Journal of Life Cycle Assessment*, 1-11. doi:10.1007/s11367-013-0670-5

- Ciroth, A; Muller, S; Weidema, B.P.; Lesage, P. (2012). *Refinig the pedigree matrix approach in ecoinvent* (Version 7.1). Report for the ecoinvent centre
- Clavreul, J., Guyonnet, D., Tonini, D., & Christensen, T. (2013). Stochastic and epistemic uncertainty propagation in LCA. *The International Journal of Life Cycle Assessment*, 1-11. doi:10.1007/s11367-013-0572-6
- Cooper, J., & Kahn, E. (2012). Commentary on issues in data quality analysis in life cycle assessment. *The International Journal of Life Cycle Assessment*, 17(4), 499-503. doi:10.1007/s11367-011-0371-x
- Coulon, R., Camobreco, V., Teulon, H., & Besnainou, J. (1997). Data quality and uncertainty in LCI. *The International Journal of Life Cycle Assessment*, 2(3), 178-182. Retrieved from <http://dx.doi.org/10.1007/BF02978816>
- Daniels, W. M., & Higgins, N. A. (2002). *Environmental Distributions and Practical Utilisation of Detection Limited Environment Measurement Data*. Retrieved from Oxon:
- Ecoinvent. (2010). Database ecoinvent data v2.2 Retrieved from <http://www.ecoinvent.org/database/>. from Swiss center of life cycle inventories <http://www.ecoinvent.org/database/>
- Ecoinvent (Producer). (2014, 2015-10-23). activity overview for ecoinvent database. [Excel spreadsheets] Retrieved from https://www.ecoinvent.org/files/activity_overview_for_users_3.1_default.xlsx
- European Commission, Joint Research Centre, & Institute for Environment and Sustainability. (2010). *International Reference Life Cycle Data System (ILCD) Handbook - Specific guide for Life Cycle Inventory data sets*. Retrieved from Luxembourg:
- European Commission, Joint Research Centre & Institute for Environment and Sustainability,. (2010a). *Framework and requirements for Life Cycle Impact Assesement models and indicators*. Retrieved from Ispra: <http://lct.jrc.ec.europa.eu/pdf-directory/ILCD-Handbook-LCIA-Framework-requirements-online-12March2010.pdf>
- European Commission, Joint Research Centre & Institute for Environment and Sustainability,. (2010b). *International Reference Life Cycle Data System (ILCD) Handbook - Specific guide for Life Cycle Inventory data sets*. Retrieved from Luxembourg:
- Finnveden, G., Hauschild, M. Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., Suh, S. (2009). Recent developments in Life Cycle Assessment. *Journal of Environmental Management*, 91(1), 1-21. doi:10.1016/j.jenvman.2009.06.018
- Frischknecht, R., Jungbluth, N., Althaus, H.-J., Doka, G., Dones, R., Heck, T., Spielmann, M. (2005). The ecoinvent Database: Overview and Methodological Framework (7 pp). *The International Journal of Life Cycle Assessment*, 10(1), 3-9. doi:10.1065/lca2004.10.181.1
- Funtowicz, S. O., & Ravetz, J. R. (1990). *Uncertainty and quality in science for policy*. Dordrecht: Kluwer Academic Publishers.
- GaBi. (2007). *GaBi Modelling Principles*. Retrieved from Leinfelden-Echterdingen:
- Geisler, G., Hellweg, S., & Hungerbühler, K. (2005). Uncertainty Analysis in Life Cycle Assessment (LCA): Case Study on Plant-Protection Products and Implications for

- Decision Making (9 pp + 3 pp). *The International Journal of Life Cycle Assessment*, 10(3), 184-192. doi:10.1065/lca2004.09.178
- Gill, J. W., Lee D. (2005). Elicited Priors for Bayesian Model Specifications in Political Science Research. *The Journal of Politics*, 67(3).
- Grabowski, A., Selke, S. M., Auras, R., Patel, M., & Narayan, R. (2015). Life cycle inventory data quality issues for bioplastics feedstocks. *The International Journal of Life Cycle Assessment*, 20(5), 584-596. doi:10.1007/s11367-015-0853-3
- Gouvernement du Québec, Loi sur la qualité de l'environnement, 53.4.1 C.F.R. (2011).
- Grivel, L.; Guillemin-Lanne, S.; Coupet, P.; Huot, C. (2001). *Analyse en ligne de l'information: une approche permettant l'extraction d'informations stratégiques basée sur la construction de connaissances*. Retrieved from Paris:
- Hamid, H. A., Yahaya, A. S., Ramli, N. A., & Ul-Saufie, A. Z. (2013). Finding the Best Statistical Distribution Model in PM10 Concentration Modeling by using Lognormal Distribution. *Journal of Applied Science*, 13(2), 294-300. doi:10.3923/jas.2013.294.300
- Hansen, J. J., Olivia; Rosling, Hans. (2012, 2012). *Science is not enough*. Paper presented at the AAAS Annual Meeting, Vancouver.
- Heijungs, R., & Frischknecht, R. (2005). Representing Statistical Distributions for Uncertain Parameters in LCA. Relationships between mathematical forms, their representation in EcoSpold, and their representation in CMLCA (7 pp). *The International Journal of Life Cycle Assessment*, 10(4), 248-254. Retrieved from <http://dx.doi.org/10.1065/lca2004.09.177>
- Heijungs, R., & Lenzen, M. (2014). Error propagation methods for LCA—a comparison. *The International Journal of Life Cycle Assessment*, 1-17. doi:10.1007/s11367-014-0751-0
- Heijungs, R. & Suh, S. (Ed.) (2002). *The computational structure of life cycle assessment* (Vol. 11). Dordrecht: Kluwer Academic Publishers,.
- Henriksson, P., Guinée, J., Heijungs, R., de Koning, A., & Green, D. (2014). A protocol for horizontal averaging of unit process data—including estimates for uncertainty. *The International Journal of Life Cycle Assessment*, 19(2), 429-436. doi:10.1007/s11367-013-0647-4
- Hong, J., Shaked, S., Rosenbaum, R., & Joliet, O. (2010). Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. *The International Journal of Life Cycle Assessment*, 15(5), 499-510. Retrieved from <http://dx.doi.org/10.1007/s11367-010-0175-4>
- Huijbregts, M. (1998). Application of uncertainty and variability in LCA. *The International Journal of Life Cycle Assessment*, 3(5), 273-280. doi:10.1007/bf02979835
- Huijbregts, M., Norris, G., Bretz, R., Ciroth, A., Maurice, B., von Bahr, B., de Beaufort, A. (2001). Framework for modelling data uncertainty in life cycle inventories. *The International Journal of Life Cycle Assessment*, 6(3), 127-132. doi:10.1007/bf02978728
- Huijbregts, M. A. J., Gilijamse, W., Ragas, A. M. J., & Reijnders, L. (2003). Evaluating Uncertainty in Environmental Life-Cycle Assessment. A Case Study Comparing Two

- Insulation Options for a Dutch One-Family Dwelling. *Environmental Science & Technology*, 37(11), 2600-2608. doi:10.1021/es020971+
- Imbeault-Tétrault, H. (2010). *Propagation analytique de l'incertitude à travers le calcul matriciel d'une analyse de cycle de vie*. (Maîtrise ès sciences appliquées), École Polytechnique de Montréal, Montréal.
- Imbeault-Tétrault, H., Jolliet, O., Deschênes, L., & Rosenbaum, R. K. (2013). Analytical Propagation of Uncertainty in Life Cycle Assessment Using Matrix Formulation. *Journal of Industrial Ecology*, n/a-n/a. doi:10.1111/jiec.12001
- INIES. (2013). Base nationale de référence sur les impacts environnementaux et sanitaires des produits, équipements et services pour l'évaluation de la performance des ouvrages. France: INIES.
- ISO. (2005). ISO 9000:2005 Systèmes de management de la qualité - Principes essentiels et vocabulaires *Principes de management de la qualité*. Genève: International Standardization Organisation.
- ISO. (2006a). ISO 14040:2006 Environmental management - Life cycle assessment - Principles and framework. Suisse: ISO.
- ISO. (2006b). ISO 14044:2006 Environmental management-Life cycle assessment-Requirements and guidelines (First ed.). Suisse: ISO.
- Jager, T. (1998). *Uncertainty analysis of EUSES: interviews with representatives from Member States and industry*. Retrieved from Bilthoven, the Netherlands:
- Jolliet, O. Saadé, M; Crettaz, Pierre (Ed.) (2005). *Analyse du cycle de vie - Comprendre et réaliser un écobilan*. Lausanne.
- Junnila, S., & Horvath, A. (2003). Life-Cycle Environmental Effects of an Office Building. *Journal of Infrastructure Systems*, 9(4), 157-166. Retrieved from [http://dx.doi.org/10.1061/\(ASCE\)1076-0342\(2003\)9:4\(157\)](http://dx.doi.org/10.1061/(ASCE)1076-0342(2003)9:4(157))
- Karlis, D.; Xekalaki, E. (2008). *Advanced Mathematical and Statistical Modeling* Statistics for industry and technology (pp. 368). doi:10.1007/978-0-8176-4626-4
- Katz, R. W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate research*, 20, 167-185.
- Kennedy, D., Montgomery, D., & Quay, B. (1996). Data quality: Stochastic environmental life cycle assessment modeling. *The International Journal of Life Cycle Assessment*, 1(4), 199-207. Retrieved from <http://dx.doi.org/10.1007/BF02978693>
- Keynes, J. M. (1937). *La théorie générale de l'emploi*.
- Koffler, C., & Kalish, D. (2012). *Quick-but-Clean? A Screening Tool using BERT Estimates to Incorporate Uncertainty*. Paper presented at the LCA XII, Tacoma (WA).
- Larousse. (Ed.) (1995a) *Le petit Larousse illustré - Dictionnaire encyclopédique*. Paris.
- Larousse. (Ed.) (1995b) *Le petit Larousse illustré - Dictionnaire encyclopédique*. Paris.
- Lélé, S. M. (1991). Sustainable development: A critical review. *World Development*, 19(6), 607-621. doi:10.1016/0305-750X(91)90197-P

- Leroy, Y. (2009). *Développement d'une méthodologie de fiabilisation des prises de décisions environnementales dans le cadre d'analyses du cycle de vie basée sur l'analyse et la gestion des incertitudes sur les données d'inventaires*. (Docteur), Ecole Nationale Supérieure d'Arts et Métiers, Paris.
- Lewandowska, A., Foltynowicz, Z., & Podlesny, A. (2004). Comparative lca of industrial objects part 1: LCA data quality assurance — sensitivity analysis and pedigree matrix. *The International Journal of Life Cycle Assessment*, 9(2), 86-89. Retrieved from <http://dx.doi.org/10.1007/BF02978567>
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal Distributions across the Science: Keys and Clues. *BioScience*, 51(5), 341-351.
- Lloyd, S. M., & Ries, R. (2007). Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches. *Journal of Industrial Ecology*, 11(1), 161-179. Retrieved from <http://dx.doi.org/10.1162/jiec.2007.1136>
- Lo, S.-C., Ma, H.-w., & Lo, S.-L. (2005). Quantifying and reducing uncertainty in life cycle assessment using the Bayesian Monte Carlo method. *Science of The Total Environment*, 340(1-3), 23-33. doi:10.1016/j.scitotenv.2004.08.020
- MacLeod, M., Fraser, A. J., & Mackay, D. (2002). Evaluating and expressing the propagation of uncertainty in chemical fate and bioaccumulation models. *Environmental Toxicology and Chemistry*, 21(4), 700-709. Retrieved from <http://dx.doi.org/10.1002/etc.5620210403>
- Manneh, R. (2011). *Évaluation de l'incertitude lors du développement de la fraction prise pour la détermination des impacts sur la santé humaine dans le cadre de l'analyse du cycle de vie*. (Ph.D.), Université de Montréal, Montréal.
- Mastrandrea, M., Mach, K., Plattner, G.-K., Edenhofer, O., Stocker, T., Field, C., Matschoss, P. (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Climatic Change*, 108(4), 675-691. doi:10.1007/s10584-011-0178-6
- Maurice, B., Frischknecht, R., Coelho-Schwartz, V., & Hungerbühler, K. (2000). Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. *Journal of Cleaner Production*, 8(2), 95-108. doi:10.1016/S0959-6526(99)00324-8
- Miller, S. A., Moysey, S., Sharp, B., & Alfaro, J. (2013). A Stochastic Approach to Model Dynamic Systems in Life Cycle Assessment. *Journal of Industrial Ecology*, 17(3), 352-362. doi:10.1111/j.1530-9290.2012.00531.x
- Morgan, M. G., Henrion, M., & Small, M. (1992). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*: Cambridge University Press.
- Morita, S., Thall, P., & Müller, P. (2010). Evaluating the Impact of Prior Assumptions in Bayesian Biostatistics. *Statistics in Biosciences*, 2(1), 1-17. doi:10.1007/s12561-010-9018-x
- Muller, S., Lesage, P., Ciroth, A., Mutel, C., Weidema, B., & Samson, R. (2014). The application of the pedigree approach to the distributions foreseen in ecoinvent v3. *The International Journal of Life Cycle Assessment*, 1-11. doi:10.1007/s11367-014-0759-5

- Mullins, K. A., Griffin, W. M., & Matthews, H. S. (2010). Policy Implications of Uncertainty in Modeled Life-Cycle Greenhouse Gas Emissions of Biofuels. *Environmental Science & Technology*, 45(1), 132-138. doi:10.1021/es1024993
- Narita, N., Nakahara, Y., Morimoto, M., Aoki, R., & Suda, S. (2004). Current LCA database development in Japan -results of the LCA project. *The International Journal of Life Cycle Assessment*, 9(6), 355-359. doi:10.1007/bf02979077
- National Renewable Energy Laboratory, A. S. M. I., Franklin Associates Ltd, Sylvatica. (2004). *U.S. LCI Database Project - User's Guide*. Retrieved from Golden:
- O'Hagan, A.; Buck, E., C.; Daneshkhan, A.; Eiser, J.R.; Garthwaite, P.H. ; Jenkinson, D.J. ; Oakley, J.E.; Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities* (Wiley Ed.): Wiley.
- Pizzol, M., Christensen, P., Schmidt, J., & Thomsen, M. (2011a). Eco-toxicological impact of “metals” on the aquatic and terrestrial ecosystem: A comparison between eight different methodologies for Life Cycle Impact Assessment (LCIA). *Journal of Cleaner Production*, 19(6-7), 687-698. doi:10.1016/j.jclepro.2010.12.008
- Pizzol, M., Christensen, P., Schmidt, J., & Thomsen, M. (2011b). Impacts of “metals” on human health: a comparison between nine different methodologies for Life Cycle Impact Assessment (LCIA). *Journal of Cleaner Production*, 19(6-7), 646-656. doi:http://dx.doi.org/10.1016/j.jclepro.2010.05.007
- Qian, S. S., Stow, C. A., & Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. *Ecological Modelling*, 159(2-3), 269-277. doi:10.1016/s0304-3800(02)00299-5
- Reap, J., Roman, F., Duncan, S., & Bras, B. (2008). A survey of unresolved problems in life cycle assessment. *The International Journal of Life Cycle Assessment*, 13(5), 374-388. doi:10.1007/s11367-008-0009-9
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software*, 22(11), 1543-1556. doi:10.1016/j.envsoft.2007.02.004
- Röös, E., Sundberg, C., & Hansson, P.-A. (2010). Uncertainties in the carbon footprint of food products: a case study on table potatoes. *The International Journal of Life Cycle Assessment*, 15(5), 478-488. doi:10.1007/s11367-010-0171-8
- Rousseaux, P., Labouze, E., Suh, Y., Blanc, I., Gaveglia, V., & Navarro, A. (2001). An overall assessment of Life Cycle Inventory quality. *The International Journal of Life Cycle Assessment*, 6(5), 299-306. Retrieved from http://dx.doi.org/10.1007/BF02978792
- Roy, P.-O. (2012). *Modélisation environnementale régionalisée à l'échelle mondiale de l'acidification terrestre et aquatique dans le cadre de l'évaluation des impacts du cycle de vie*. (Philosophiae Doctor), Polytechnique Montréal, Montréal.
- Silva, C. (2011). Electric and plug-in hybrid vehicles influence on CO₂ and water vapour emissions. *International Journal of Hydrogen Energy*, 36(20), 13225-13232. doi:http://dx.doi.org/10.1016/j.ijhydene.2011.07.023

- Slob, W. (1994). Uncertainty Analysis in Multiplicative Models. *Risk Analysis*, 14(4), 571-576. Retrieved from <http://dx.doi.org/10.1111/j.1539-6924.1994.tb00271.x>
- Sonnemann, G. W., Schuhmacher, M., & Castells, F. (2003). Uncertainty assessment by a Monte Carlo simulation in a life cycle inventory of electricity produced by a waste incinerator. *Journal of Cleaner Production*, 11(3), 279-292. doi:10.1016/s0959-6526(02)00028-8
- Steinmann, Z. J. N., Venkatesh, A., Hauck, M., Schipper, A. M., Karuppiyah, R., Laurenzi, I. J., & Huijbregts, M. A. J. (2014). How To Address Data Gaps in Life Cycle Inventories: A Case Study on Estimating CO₂ Emissions from Coal-Fired Electricity Plants on a Global Scale. *Environmental Science & Technology*, 48(9), 5282-5289. doi:10.1021/es500757p
- Sugiyama, H., Fukushima, Y., Hirao, M., Hellweg, S., & Hungerbühler, K. (2005). Using Standard Statistics to Consider Uncertainty in Industry-Based Life Cycle Inventory Databases (7 pp). *The International Journal of Life Cycle Assessment*, 10(6), 399-405. doi:10.1065/lca2005.05.211
- Swart, R., Bernstein, L., Ha-Duong, M., & Petersen, A. (2009). Agreeing to disagree: uncertainty management in assessing climate change, impacts and responses by the IPCC. *Climatic Change*, 92(1), 1-29. doi:10.1007/s10584-008-9444-7
- Tan, R. R., Briones, L. M. A., & Culaba, A. B. (2007). Fuzzy data reconciliation in reacting and non-reacting process data for life cycle inventory analysis. *Journal of Cleaner Production*, 15(10), 944-949. doi:10.1016/j.jclepro.2005.09.001
- Tan, R. R., Culaba, A. B., & Purvis, M. R. I. (2002). Application of possibility theory in the life-cycle inventory assessment of biofuels. *International Journal of Energy Research*, 26(8), 737-745. Retrieved from <http://dx.doi.org/10.1002/er.812>
- Ukidwe, N., Bakshi, B. R., & Parthasarathy, G. (2004). *A multiscale Bayesian framework for designing efficient and sustainable industrial systems*. Paper presented at the AIChE sustainability engineering conference proceedings, Austin, TX.
- UN, Statistics Division (Producer). (2011). ISIC Rev.4 - NAICS(US) 2012, correspondences. [Zip files] Retrieved from <http://unstats.un.org/unsd/cr/registry/regdntransfer.asp?f=273>
- UNEP. (2009). *Guidelines for Social Life Cycle Assessment of Products*. Retrieved from http://www.unep.org/publications/search/pub_details_s.asp?ID=4102
- UNEP. (2015). *Our planet - Time for Global Action*. Retrieved from Nairobi, Kenya: <http://apps.unep.org/publications/index.php?ftype=Our+Planet>
- van der Sluijs, J.; Janssen, P.H.M.; Petersen, A.C.; Klopogge, P.; Risbey, J.S.; Tuinstra, W. ; Ravetz, J.R. (2004). *RIVM/MNP Guidance for Uncertainty Assessment and Communication: Tool Catalogue for Uncertainty Assessment* (NWS-E-2004-37). Retrieved from Utrecht: <http://www.nusap.net/downloads/toolcatalogue.pdf>
- van der Sluijs, J. P., Craye, M., Funtowicz, S., Klopogge, P., Ravetz, J., & Risbey, J. (2005). Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, 25(2), 481-492. doi:10.1111/j.1539-6924.2005.00604.x
- Venkatesh, A., Jaramillo, P., Griffin, W. M., & Matthews, H. S. (2010). Uncertainty Analysis of Life Cycle Greenhouse Gas Emissions from Petroleum-Based Fuels and Impacts on Low

- Carbon Fuel Policies. *Environmental Science & Technology*, 45(1), 125-131. doi:10.1021/es102498a
- Vrancken, D. (2014). *Penser l'incertain*. Québec: Presse de l'Université Laval.
- Wang, E., & Shen, Z. (2013). A hybrid Data Quality Indicator and statistical method for improving uncertainty analysis in LCA of complex system – application to the whole-building embodied energy analysis. *Journal of Cleaner Production*, 43(0), 166-173. doi:http://dx.doi.org/10.1016/j.jclepro.2012.12.010
- Wang, E., Shen, Z., Neal, J., Shi, J., Berryman, C., & Schwer, A. (2012). An AHP-weighted aggregated data quality indicator (AWADQI) approach for estimating embodied energy of building materials. *The International Journal of Life Cycle Assessment*, 1-10. doi:10.1007/s11367-012-0417-8
- Wardekker, J. A., van der Sluijs, J. P., Janssen, P. H. M., Klopogge, P., & Petersen, A. C. (2008). Uncertainty communication in environmental assessments: views from the Dutch science-policy interface. *Environmental Science & Policy*, 11(7), 627-641. doi:10.1016/j.envsci.2008.05.005
- Warmink, J., Klis, H., Booij, M., & Hulscher, S. M. H. (2011). Identification and Quantification of Uncertainties in a Hydrodynamic River Model Using Expert Opinions. *Water Resources Management*, 25(2), 601-622. doi:10.1007/s11269-010-9716-7
- Weber, C. L. (2012). Uncertainty and Variability in Product Carbon Footprinting. *Journal of Industrial Ecology*, no-no. doi:10.1111/j.1530-9290.2011.00407.x
- Weidema, B. P. (1998). Multi-user test of the data quality matrix for product life cycle inventory data. *The International Journal of Life Cycle Assessment*, 3(5). doi:10.1007/BF02979832
- Weidema, B. P., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Vadenbo, C. O., & Wernet, G. (2013). *Overview and methodology. Data quality guidelines for the Ecoinvent database version 3. Ecoinvent Report 1 (final)*. Retrieved from St Gallen:
- Weidema, B. P., & Wesnæs, M. S. (1996). Data quality management for life cycle inventories--an example of using data quality indicators. *Journal of Cleaner Production*, 4(3-4), 167-174. doi:Doi: 10.1016/s0959-6526(96)00043-1
- Weidema, B. P.; Bauer., C. ; Hischier,R. ; Mutel,C. ;Nemecek,T. ;Vadenbo, C.O. ; Wernet,G. (2013). *Overview and methodology. Data quality guidelines for the Ecoinvent database version 3. Ecoinvent Report 1 (final)*. Retrieved from St Gallen: ecoinvent centre
- Weidema, B. P.; Cappellaro., F. ; Carlson,R. ; Notten,P. ; Palsson,A.C. ; Patyk,A. ; Regalini,E. ; Sacchetto,F. ; Scalbi,S. (2003). *Procedural guideline for collection, treatment, and quality documentation of LCA data*. Retrieved from 2.0 LCA Consultants
- Williams, E. D., Weber, C. L., & Hawkins, T. R. (2009). Hybrid Framework for Managing Uncertainty in Life Cycle Inventories. *Journal of Industrial Ecology*, 13(6), 928-944. doi:10.1111/j.1530-9290.2009.00170.x

ANNEXE A – DES DONNÉES AUTONOMES AUX IMPACTS POTENTIELS, LE MODÈLE ACV

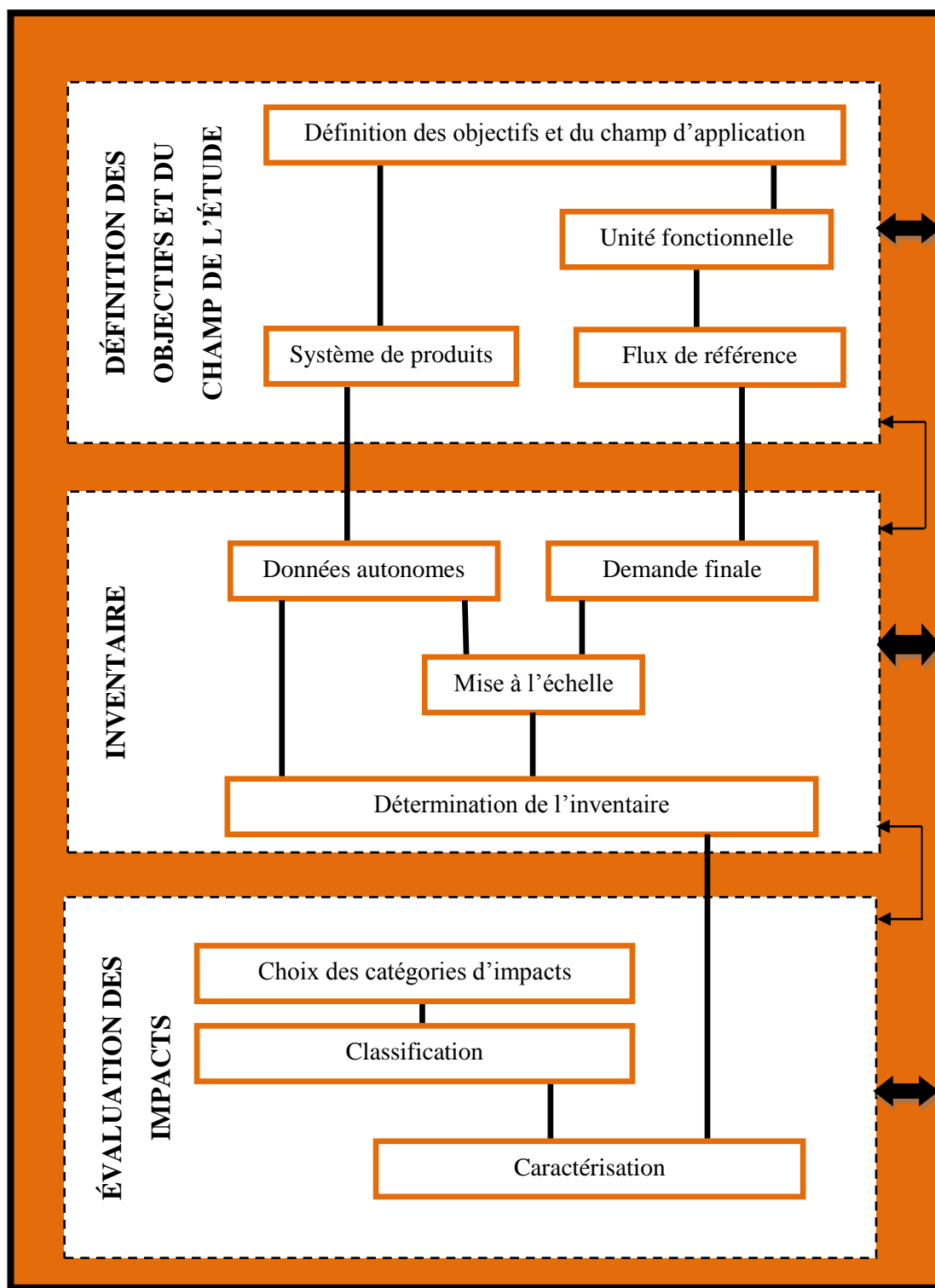


Figure 9-1 : Schéma méthodologique général de mise en œuvre d'une ACV – adapté de ISO 14040 (ISO, 2006a)

Afin de passer de l'unité fonctionnelle à la quantification des impacts environnementaux potentiels d'un système à l'étude, un certain nombre d'étapes sont à accomplir, la plupart basées sur des modèles de calcul. La Figure 9-1 adaptée de la norme ISO 14040 se veut un recensement non exhaustif de certaines des étapes à accomplir lors de la réalisation d'une ACV. Seules les étapes nécessaires à la compréhension globale sont présentées à cette figure et font l'objet de la présentation ci-dessous.

La **phase de définition des objectifs et du champ de l'étude** donne le contexte de l'étude et son domaine d'application. Une fois la fonction du système quantifiée par le biais de *l'unité fonctionnelle*, il s'agit de définir les *flux de références* qui sont les quantités de produits nécessaires pour remplir l'unité fonctionnelle. Un autre élément qu'il est nécessaire de définir est le *système* (ou l'arbre) *de produit*. Celui-ci est un agencement de *processus élémentaires* qui permet de modéliser au plus proche de la réalité le système à l'étude et qui définit ses *frontières*. Un arbre de produit au complet compte généralement quelques centaines de processus élémentaires tels que représentés à la Figure 1-1. On distingue deux types de flux entrant ou sortant d'un processus élémentaire : les *flux élémentaires* qui représentent les échanges entre le processus et l'environnement et les *flux économiques* qui traduisent un échange entre les processus élémentaires.

La **phase d'inventaire** consiste à quantifier l'ensemble des flux élémentaires qui traversent les frontières du système, et ce, par unité de flux de référence. Afin d'effectuer cette étape, il s'agit tout d'abord de quantifier l'ensemble des flux élémentaires et économiques par processus élémentaire: ceci passe par des relevés sur le terrain, de la revue de littérature; un nombre important de processus élémentaires ont d'ores et déjà été modélisés et sont disponibles dans des bases de données d'inventaire telles qu'ecoinvent par exemple. Ces données doivent ensuite être ramenées par unité de flux de référence par une *mise à l'échelle*.

Afin de réaliser la mise à l'échelle des flux, plusieurs méthodes de calcul existent chacune avec leurs avantages et leurs inconvénients. La plupart des logiciels permettant de réaliser des ACV utilisent une approche matricielle afin de calculer l'inventaire du système au complet. Le modèle

de base de calcul est présenté ici, les modèles de calcul raffinés permettant de répondre à certaines des problématiques de l'ICV sont recensés dans l'ouvrage de Heijungs & Suh (2002).

L'ensemble de la collecte de données permet de définir la matrice technologique **A** qui regroupe la quantification de l'ensemble des flux économiques pour chacun des processus élémentaires présents dans l'arbre de produit. Les flux de référence permettent, quant à eux, de définir le vecteur de demande finale **f** regroupant l'ensemble des quantités de produits nécessaire pour remplir l'unité fonctionnelle. Afin de ramener les quantités présentées dans la matrice technologique par unité de flux de référence, un vecteur **s** de mise à l'échelle est défini tel que **As=f**.

De la même manière, l'ensemble des flux élémentaires de chaque processus sont confinés dans la matrice environnementale **B** qui une fois mise à l'échelle (i.e. ramenée aux flux de référence) va permettre d'avoir accès au vecteur regroupant l'ensemble des échanges entre le système et l'environnement : le vecteur d'inventaire **g**. On a donc **Bs=g**. La première équation matricielle permet d'avoir accès au vecteur **s** par inversion de la matrice **A** – dans le cas le plus simple, la matrice **A** est carrée et non singulière, se référer à l'ouvrage d'Heijungs et Suh pour les cas particuliers -. Le vecteur **g** s'obtient alors par **g=BA⁻¹f**.

La **phase d'évaluation des impacts** permet de caractériser le vecteur d'inventaire **g** afin d'obtenir les impacts environnementaux potentiels du système à l'étude. Afin d'obtenir les facteurs de caractérisation, les développeurs se basent sur des modélisations environnementales, toxicologiques et éco-toxicologiques qui ne sont pas précisées ici. Ces facteurs se retrouvent dans une matrice de caractérisation **C** qui multipliée par le vecteur d'inventaire va permettre d'obtenir le vecteur des scores d'impact **h** (**h=Cg**).

Si la **phase d'interprétation** permet de conclure quant aux objectifs de l'analyse, l'ACV étant un processus itératif, cette phase a cours tout au long du modèle et permet de prendre des décisions lors de la réalisation de l'étude.

La prise en compte de l'incertitude dans le modèle matriciel

L'approche matricielle permet de considérer la propagation des incertitudes que ce soit par la méthode analytique ou encore par approche de Monte Carlo (cf. section 2.3.3) (Heijungs & Suh, 2002).

L'incertitude peut être quantifiée sur chacun des éléments des matrices technologique, environnementale ou encore de caractérisation ainsi que sur le vecteur de demande finale. Néanmoins, dans le cas de la simulation de Monte Carlo, approche largement utilisée dans les logiciels d'ACV, la quantification de l'incertitude sur le vecteur **h** des scores d'impact est une étape longue. Ce temps de calcul important est une des raisons pour lesquelles les analyses quantitatives d'incertitude ne sont pas monnaie courante dans les études d'ACV.

ANNEXE B – MATÉRIEL SUPPLÉMENTAIRE PUBLIÉ AVEC L'ARTICLE 1 (CHAPITRE 4)

Les éléments suivants ont été publiés en matériel supplémentaire pour le premier article présenté au Chapitre 4. Compte tenu de la nature des résultats des ratios sur les paramètres statistiques de chacune des données (3981 systèmes de produits x 16 ratios), ces derniers ne sont pas représentés ici. Seuls les systèmes de produits utilisés pour la comparaison ainsi que les résultats complémentaires associés à cette comparaison sont présentés dans cette annexe.

Systèmes de produits utilisés pour la comparaison

- voir pages suivantes -

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
2277	Stirling cogen unit 3kWe, wood pellets, future/CH/I U	energy	Biomass	Cogeneration	Infrastructure
2319	Electricity, wood, at distillery/CH U	energy	Biomass	Power plants	
831	Catalytic converter, three-way, Mini CHP plant/CH/I U	energy	Cogeneration	Gas	Power unit
828	Catalytic converter, SCR, 200 litre/RER/I U	energy	Cogeneration	Oil	Power unit
510	Cogen unit 6400kWth, wood burning, common components for heat+electricity/CH/I U	energy	Cogeneration	Wood	Power unit
1938	Hard coal, burned in power plant/NL U	energy	Electricity by fuel	Coal	Power plant
3277	Natural gas, burned in gas turbine, for compressor station/RU U	energy	Electricity by fuel	Gas	Gas turbine
2186	Natural gas, burned in power plant/BE U	energy	Electricity by fuel	Gas	Power plant
1635	Electricity, hydropower, at run-of-river power plant/CH U	energy	Electricity by fuel	Hydro	Flow through
1771	Electricity, hydropower, at pumped storage power plant/HR U	energy	Electricity by fuel	Hydro	Pumping storage
1636	Electricity, hydropower, at reservoir power plant/CH U	energy	Electricity by fuel	Hydro	Reservoir
3056	Lignite power plant/RER/I U	energy	Electricity by fuel	Lignite	Power plant
2228	Nuclear power plant, boiling water reactor 1000MW/DE/I U	energy	Electricity by fuel	Nuclear	BWR
2243	Electricity, nuclear, at power plant pressure water reactor/DE U	energy	Electricity by fuel	Nuclear	PWR
2257	Heavy fuel oil, burned in power plant/CS U	energy	Electricity by fuel	Oil	Power plant
2275	Peat, burned in power plant/NORDEL U	energy	Electricity by fuel	Peat	Power plant
94	3kWp facade installation, multi-Si, panel, mounted, at building/CH/I U	energy	Electricity by fuel	Photovoltaic	Infrastructure
1920	Electricity, digester sludge, at incineration plant, future, alloc. price/CH U	energy	Electricity by fuel	Waste	
1900	Wind power plant 30kW, fixed parts/CH/I U	energy	Electricity by fuel	Wind	Infrastructure
2322	Electricity, wood, at distillery/SE U	energy	Electricity by fuel	Wood	
2005	Electricity, high voltage, production GB, at grid/GB U	energy	Electricity country mix	High Voltage	
1953	Electricity, high voltage, aluminium industry, at grid/GLO U	energy	Electricity country mix	High Voltage + import	
2111	Electricity, low voltage, consumer mix, at grid/CH U	energy	Electricity country mix	Low Voltage	
2063	Electricity, low voltage, at grid/BG U	energy	Electricity country mix	Low Voltage + import	
2130	Electricity, medium voltage, production CZ, at grid/CZ U	energy	Electricity country mix	Medium Voltage	
2106	Electricity, medium voltage, at grid/SI U	energy	Electricity country mix	Medium Voltage + import	
1712	Electricity, production mix AT/AT U	energy	Electricity country mix	Production	
1831	Electricity mix/US U	energy	Electricity	Production +	

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
			country mix	import	
2688	Industrial furnace, coal, 1-10 MW/RER/I U	energy	Heat	Coal	Furnace
65	Natural gas, burned in industrial furnace >100kW/RER U	energy	Heat	Gas	Furnace
717	Borehole heat exchanger 150 m/CH/I U	energy	Heat	Heat pump	Infrastructure
2513	Lignite briquette, burned in stove 5-15kW/RER U	energy	Heat	Lignite	Furnace
2887	Industrial furnace 1MW, oil/CH/I U	energy	Heat	Oil	Furnace
40	Heat, unspecific, in chemical plant/RER U	energy	Heat	Others	
2758	Solar system, flat plate collector, one-family house, combined system/CH/I U	energy	Heat	Solar	Infrastructure
791	Steam, for chemical processes, at plant/RER U	energy	Heat	Steam	
2807	Heat, digester sludge, at incineration plant, future, allocation price/CH U	energy	Heat	Waste	
2533	Furnace, logs, hardwood, 6kW/CH/I U	energy	Heat	Wood	Furnace
1272	Diesel-electric generating set production 10MW/RER/I U	energy	Mechanical	Infrastructure	
2426	Facade construction, integrated, at building/CH/I U	energy	Others	Obsolete	
2634	Grass silage organic, at farm/CH U	material	Agricultural	Animal production	Animal foods
2366	Molasses, from sugar beet, at sugar refinery/CH U	material	Agricultural	Food	Byproducts
508	Sugar refinery/GLO/I U	material	Agricultural	Food	Infrastructure
661	Biowaste, at collection point/CH U	material	Agricultural	Others	
3666	Rape meal, at oil mill/CH U	material	Agricultural	Plant oils	Byproducts
3340	Oil mill/CH/I U	material	Agricultural	Plant oils	Infrastructure
2440	Pea seed IP, at regional storehouse/CH U	material	Agricultural	Plant production	Seeds
863	Ceramic plant/CH/I U	material	Ceramics	Infrastructure	
2494	Phosphoric acid plant, fertiliser grade/US/I U	material	Chemicals	Acids (inorganic)	Infrastructure
1227	Formic acid, at plant/RER U	material	Chemicals	Acids (organic)	
2502	Potassium nitrate, as N, at regional storehouse/RER U	material	Chemicals	Fertilisers (inorganic)	
1019	Compost plant, open/CH/I U	material	Chemicals	Fertilisers (organic)	Infrastructure
250	Air separation plant/RER/I U	material	Chemicals	Gases	Infrastructure
392	Anhydrite, burned, at plant/CH U	material	Chemicals	Inorganic	
41	Chemical plant, organics/RER/I U	material	Chemicals	Organic	Infrastructure
252	Facilities, chemical production/RER/I U	material	Chemicals	Others	Infrastructure
4145	Wood preservative, creosote, at plant/RER U	material	Chemicals	Pesticides	
1234	Silicone plant/RER/I U	material	Chemicals	Silicons	Infrastructure
1237	DAS-1, fluorescent whitening agent triazinylaminostilben type, at	material	Chemicals	Washing agents	Auxiliaries

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
	plant/RER U				
1254	Sodium metasilicate pentahydrate, 58%, powder, at plant/RER U	material	Chemicals	Washing agents	Builders
2387	Ethoxylated alcohols (AE11), palm oil, at plant/RER U	material	Chemicals	Washing agents	Tensides
704	Cement plant/CH/I U	material	Construction	Binders	Infrastructure
381	Bitumen, at refinery/CH U	material	Construction	Bitumen	
707	Refractory, basic, packed, at plant/DE U	material	Construction	Bricks	
933	Cladding, crossbar-pole, aluminium, at plant/RER U	material	Construction	Cladding	
200	Concrete mixing plant/CH/I U	material	Construction	Concrete	Infrastructure
733	Fibre cement corrugated slab, at plant/CH U	material	Construction	Coverings	
1590	Door, inner, wood, at plant/RER U	material	Construction	Doors	
2499	Foam glass plant/BE/I U	material	Construction	Insulation	Infrastructure
263	Building, hall/CH/I U	material	Construction	Others	Infrastructure
224	Alkyd paint, white, 60% in solvent, at plant/RER U	material	Construction	Paints	
665	Bitumen adhesive compound, cold, at plant/RER U	material	Construction	Sealing	
245	Air filter, decentralized unit, 180-250 m ³ /h, at plant/RER U	material	Construction	Ventilation	Components
4136	Window frame, aluminium, U=1.6 W/m ² K, at plant/RER U	material	Construction	Windows	
456	Electronic component production plant/GLO/I U	material	Electronics	Devices	Module
138	Electronics for control units/RER U	material	Electronics	Others	
117	Photovoltaic panel, CIS, at plant/DE/I U	material	Electronics	Photovoltaic	Infrastructure
2621	Grass fibres, at digestion/CH U	material	Fuels	Biofuels	Biogas
2365	Ethanol, 95% in H ₂ O, from sugar beet molasses, at distillery/CH U	material	Fuels	Biofuels	Ethanol
3163	Synthetic gas plant/CH/I U	material	Fuels	Biofuels	Infrastructure
3152	Methane, 96 vol-%, from biogas, at purification/CH U	material	Fuels	Biofuels	Methane
2675	Hard coal supply mix/PT U	material	Fuels	Coal	Coal by country
2647	Hard coal briquettes, at plant/RER U	material	Fuels	Coal	Cokes/briquets
2650	Hard coal, at regional storage/AU U	material	Fuels	Coal	Produced coal
3054	Lignite, at mine/RER U	material	Fuels	Lignite	Mine
3242	Natural gas, at long-distance pipeline/CH U	material	Fuels	Natural gas	Gas by country
59	Natural gas, high pressure, at consumer/RER U	material	Fuels	Natural gas	Gas to user
3166	Natural gas, at production onshore/DZ U	material	Fuels	Natural gas	Produced gas
404	Petroleum coke, at refinery/RER U	material	Fuels	Oil	Coke
998	Coke oven gas, at plant/DE U	material	Fuels	Oil	Coke oven gas
689	Crude oil, production RU, at long distance transport/RER U	material	Fuels	Oil	Crude oil
440	Diesel, at regional storage/CH U	material	Fuels	Oil	Diesel

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
3043	Light fuel oil, at refinery/RER U	material	Fuels	Oil	Fuel oil
1277	Regional distribution, oil products/RER/I U	material	Fuels	Oil	Infrastructure
2553	Kerosene, at regional storage/RER U	material	Fuels	Oil	Kerosene
163	Naphtha, APME mix, at refinery/RER U	material	Fuels	Oil	Naphtha
3518	Petrol, 15% vol. ETBE additive, with ethanol from biomass, at refinery/RER U	material	Fuels	Oil	Petrol
767	Propane/ butane, at refinery/RER U	material	Fuels	Oil	Propane/butane
3696	Refinery gas, at refinery/RER U	material	Fuels	Oil	Refinery gas
3515	Peat, at mine/NORDEL U	material	Fuels	Peat	
3988	Uranium enriched 3.8%, for boiling water reactor/CH U	material	Fuels	Uranium	Enriched
2247	Fuel elements PWR, UO2 3.8% & MOX, at nuclear fuel fabrication plant/FR U	material	Fuels	Uranium	Fuel element
428	Glass etching plant/DK/I U	material	Glass	Construction	Infrastructure
716	Glass production site/RER/I U	material	Glass	Packaging	Infrastructure
2567	Glass sorting site/RER/I U	material	Glass	Waste glass	Infrastructure
485	Brazing solder, cadmium free, at plant/RER U	material	Metals	Alloys	
3502	Parkes process crust, from desilverising of lead/GLO U	material	Metals	Extraction	
709	Blast oxygen furnace converter/RER/I U	material	Metals	Ferro	Production
595	Facilities precious metal refinery/SE/I U	material	Metals	Non ferro	Production
819	Iron scrap, at plant/RER U	material	Metals	Waste metals	
923	Non-ferrous metal mine, underground/GLO/I U	material	Minerals	Infrastructure	
920	Chromite, ore concentrate, at beneficiation/GLO U	material	Minerals		
246	Solid unbleached board, SUB, at plant/RER U	material	Paper+ Board	Board	
1144	Corrugated board, fresh fibre, single wall, at plant/CH U	material	Paper+ Board	Corrugated board	
1106	Paper mill, integrated/RER/I U	material	Paper+ Board	Graphic paper	Infrastructure
795	Kraft paper, unbleached, at plant/RER U	material	Paper+ Board	Packaging paper	
713	Packaging box production unit/RER/I U	material	Paper+ Board	Packagings	Infrastructure
887	Pulp plant/RER/I U	material	Paper+ Board	Pulp	Infrastructure
4118	Waste paper sorting plant/RER/I U	material	Paper+ Board	Waste paper	Infrastructure
3575	Poly lactide, granulate, at plant/GLO U	material	Plastics	Biopolymers	
205	Acrylonitrile-butadiene-styrene copolymer, ABS, at plant/RER U	material	Plastics	Rubbers	
876	Polyvinylchloride, bulk polymerised, at plant/RER U	material	Plastics	Thermoplasts	
936	Polyurethane, rigid foam, at plant/RER U	material	Plastics	Thermosets	
3854	Textile, jute, at plant/IN U	material	Textiles		

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
3844	Water supply network/CH/I U	material	Water	Drinking water	Infrastructure
4125	Water treatment plant, deionisation/CH/I U	material	Water	Industry water	Infrastructure
1584	Wood chips, softwood, from industry, u=40%, at plant/RER U	material	Wood	Byproducts	
1113	Wooden board manufacturing plant, organic bonded boards/RER/I U	material	Wood	Products	Infrastructure
220	Agricultural machinery, general, production/CH/I U	processing	Agricultural	Operations	Infrastructure
3645	Production of carton board boxes, offset printing, at plant/CH U	processing	Cardboard		
234	Air compressor, screw-type compressor, 4 kW, at plant/RER/I U	processing	Compressed air	Generation	Infrastructure
3215	Mounting, surface mount technology, Pb-containing solder/GLO U	processing	Electronics		
425	Anti-reflex-coating, etching, solar glass/DK U	processing	Glass		
2899	Hot impact extrusion, steel, 1 stroke/RER U	processing	Metals	Chipless shaping	
3185	Milling, aluminium, large parts/RER U	processing	Metals	Chipping	
1592	Zinc coating, pieces, adjustment per um/RER U	processing	Metals	Coating	
3145	Steel product manufacturing, average metal working/RER U	processing	Metals	General manufacturing	
328	Metal working factory/RER/I U	processing	Metals	Infrastructure	
3229	Silicon carbide, recycling, at plant/RER U	processing	Metals	Others	
1623	Welding, arc, steel/RER U	processing	Metals	Welding	
605	Recultivation, bentonite mine/DE U	processing	Others		
146	Injection moulding/RER U	processing	Plastics		
2689	SOx retained, in hard coal flue gas desulphurisation/RER U	processing	Power plants		
3856	Weaving, bast fibres/IN U	processing	Textiles		
2350	Energy reduction, ventilation system, 6 x 120 m ³ /h, steel ducts, without GHE/CH U	processing	Ventilation	Energy reduction	
2328	Wastewater treatment plant, class 5/CH/I U	processing	Waste	Infrastructure	
3336	Nuclear spent fuel reprocessing plant/RER/I U	processing	Waste	Nuclear waste	Infrastructure
3610	Power saw, with catalytic converter/RER/I U	processing	Wood	Infrastructure	
253	Aircraft, freight/RER/I U	transport	Air	Operations	Infrastructure
211	Conveyor belt, at plant/RER/I U	transport	Building equipment	Infrastructure	
1573	Distribution network, electricity, low voltage/CH/I U	transport	Electricity	Infrastructure	
3159	Pipeline, natural gas, low pressure distribution network/CH/I U	transport	Pipeline	Infrastructure	
3404	Operation, maintenance, railway track, ICE/DE/I U	transport	Rail	Operations	Infrastructure
3109	Maintenance, passenger car, electric, LiMn ₂ O ₄ , city car/RER/I U	transport	Road	Operations	Infrastructure
3588	Port facilities/RER/I U	transport	Water	Operations	Infrastructure
4060	Use, computer, desktop, with LCD monitor, office use/RER U	use	Computers & network		

NUMBER	PRODUCT SYSTEM NAME	Classification level 1	Classification level 2	Classification level 3	Classification level 4
1478	Disposal, hard coal ash, 0% water, to residual material landfill/BE U	waste treatment	BELGIUM <input type="checkbox"/>		
2242	Nuclear spent fuel, in conditioning, at plant/CN U	waste treatment	CHINA <input type="checkbox"/>		
1526	Disposal, passenger car, electric, LiMn2O4, city car/RER/I U	waste treatment	Europe <input type="checkbox"/>		
1350	Disposal, cement, hydrated, 0% water, to residual material landfill/CH U	waste treatment	SWITZERLAND <input type="checkbox"/>		

Résultats complémentaires sur la comparaison

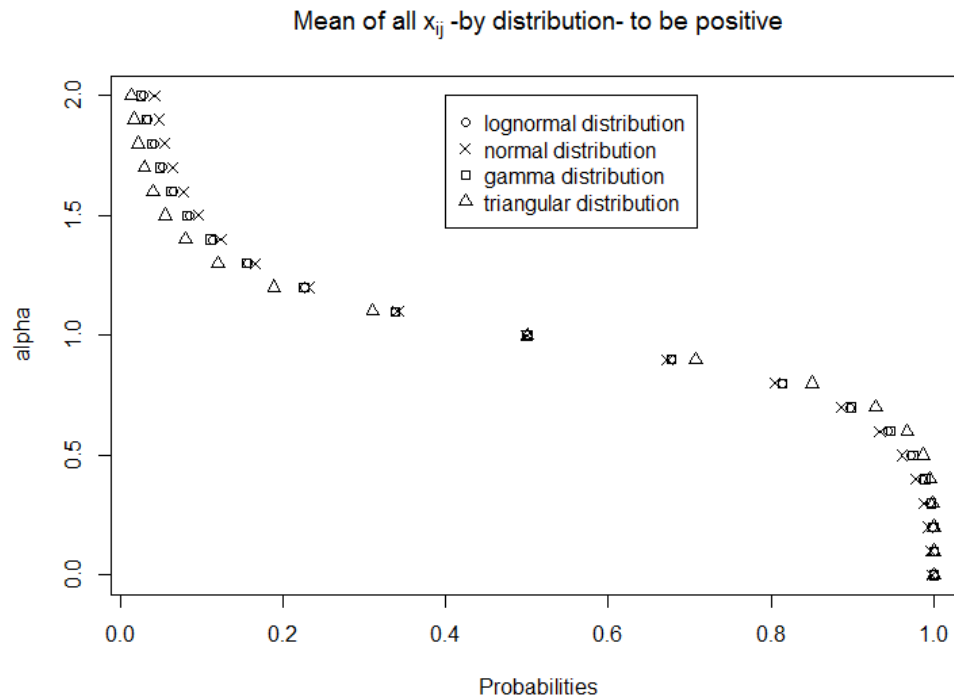
Introduction

The goal here is to investigate the effects of the switch on system comparisons when shifting the compared systems by an α -factor

Methodology

<i>Notation</i>	$P_{i,LN}$: the vector containing the 5000 outputs resulting from the Monte Carlo simulation for the i_{th} product system and the lognormal distribution. The subscripts N, G and T are respectively used for the normal, gamma and triangular distribution.
Step 1	Normalisation of the product system: $P'_{i,LN} = P_{i,LN} / \text{median}(P_{i,LN})$ (respectively for $P_{i,N}$, $P_{i,G}$, and $P_{i,T}$)
Step 2	For every couple (i,j) $x_{ij,LN} = p(P_{i,LN} - \alpha P_{j,LN} > 0)$ (respectively for $x_{ij,N}$, $x_{ij,G}$, and $x_{ij,T}$) where α is a multiplying factor that varies in the interval [0,2]. For the presented comparisons, $\alpha=0.9$
Step 3	Determination of the relative errors on the probability x_{ij} $r_{ij,LN-N} = x_{ij,N} / x_{ij,LN}$ (respectively for $r_{ij,LN-G}$ and $r_{ij,LN-T}$)
Step 4	Assessing the different relative errors obtained

Results



Conclusion

The same conclusions that the ones highlighted in the paper are shown in the figure where the mean of all x_{ij} are calculated when α varies from 0 to 2 (by a constant step of 0.1). This figure shows that the tendencies are the same, whatever distribution is used (with a slight difference when the triangular distribution is used).

ANNEXE C – MATÉRIEL SUPPLÉMENTAIRE PUBLIÉ AVEC L'ARTICLE 2 (CHAPITRE 5)

Les tableaux suivants ont été publiés en matériel supplémentaire accompagnant le deuxième article présenté dans le cadre de cette thèse.

Les tableaux suivants présentent les résultats obtenus lors des analyses de Monte Carlo conduites pour appuyer les résultats obtenus ainsi que les résultats de la comparaison entre l'approche développée dans le cadre de l'article et les résultats obtenus par analyse de Monte Carlo. Au sein de ces tableaux, l'indice $_{MC}$ se rapporte à des paramètres obtenus par les simulations de Monte Carlo alors que l'indice $_T$ se rapporte aux paramètres obtenus par les formules développés dans ce deuxième article.

Comparaison des résultats pour une loi normale

	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
σ_T	0.38	0.414	0.53	0.821
CV_T	0.253	0.276	0.353	0.547
σ_{MC}	0.38	0.42	0.56	0.93
CV_{MC}	0.252	0.277	0.366	0.554
ε_σ	0.00 %	1.43 %	5.36 %	11.72 %
ε_{CV}	0.40 %	0.36 %	3.55 %	1.26 %

Comparaison des résultats pour une loi uniforme

	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
a_T	0.991	0.921	0.677	0.039
b_T	3.009	3.079	3.323	3.961
CV_T	0.291	0.311	0.382	0.566
a_{MC}	0.93	0.75	0.42	0.2
b_{MC}	3.37	4.23	7.33	12.38
CV_{MC}	0.293	0.311	0.394	0.578
ε_a	6.16 %	18.57 %	37.96 %	80.50 %
ε_b	10.71 %	27.21 %	54.67 %	68.00 %
ε_{CV}	0.68 %	0.00 %	3.05 %	2.08 %
% of values in the interval $[a_T; b_T]$	97.10 %	94.07 %	92.15 %	90.79 %

Comparaison des résultats pour la loi triangulaire

	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
a_T	0.993	0.94	0.765	0.336
b_T	3.021	3.18	3.706	4.991
CV_T	0.234	0.254	0.314	0.435
a_{MC}	0.95	0.78	0.56	0.24
b_{MC}	3.18	3.83	4.87	12.61
CV_{MC}	0.236	0.261	0.349	0.543
ε_a	4.33 %	17.02 %	26.80 %	28.57 %
ε_b	5.00 %	16.97 %	23.90 %	60.42 %
ε_{CV}	0.85 %	2.68 %	10.03 %	19.89 %
% of values in the interval $[a_T; b_T]$	99.60 %	98.93 %	97.90 %	97.84 %

Comparaison des résultats pour la loi Beta PERT

	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
a_T	0.991	0.921	0.7	0.184
b_T	3.028	3.237	3.9	5.45
CV_T	0.235	0.255	0.322	0.464
a_{MC}	0.96	0.76	0.52	0.26
b_{MC}	2.98	3.41	5.32	10.68
CV_{MC}	0.216	0.244	0.333	0.54
ε_a	3.13 %	17.48 %	25.71 %	29.23 %
ε_b	1.59 %	5.07 %	26.69 %	48.97 %
ε_{CV}	8.09 %	4.31 %	3.30 %	14.07 %
% of values in the interval [a_T; b_T]	99.93 %	99.62 %	99.52 %	100.00 %

Comparaison des résultats pour la loi Gamma

	(2;2;2;2;2)	(3;3;3;3;3)	(4;4;4;4;4)	(5;5;5;5;5)
CV_T	0.253	0.272	0.335	0.46
CV_{MC}	0.36	0.38	0.444	0.633
ε_{CV}	29.72 %	28.42 %	24.55 %	27.33 %

ANNEXE D – MATÉRIEL SUPPLÉMENTAIRE PUBLIÉ AVEC L'ARTICLE

3 (CHAPITRE 6)

L'ensemble des tableaux suivants a été publié en matériel supplémentaire accompagnant le troisième article présenté dans le cadre de cette thèse. Il a été soumis en tant que fichier Excel.

Liste des éléments composant le « Matériel supplémentaire »

Name of sheet	Description
PedigreeMatrix	The pedigree matrix in use in ecoinvent v3
Papers	Census of all papers, reports, databases used for the development of new empirical uncertainty factors
DataClassification	Describes the way data classified for the four pedigree criteria: reliability, temporal correlation, geographical correlation and further technological correlation
AUF_1	AUF: Additional uncertainty factor The data used to derive initial estimates
AUF_2	The factors obtained in the study for all the made assumptions
AUF_3	Data used to derive additional uncertainty factors
BUF	BUF: Basic uncertainty factor Data used to derive basic uncertainty factors
Lognormal_Assumptions	Description and results for the test of the lognormal distribution assumption made for the uncertainty factors
Posterior distributions	The obtained posterior distributions for all derived uncertainty factors
Data_Gypsum	The data used to derive the results presented in section 3.3

Onglet *PedigreeMatrix*: Matrice pedigree telle qu'utilisée dans ecoinvent v3

Se référer au Tableau 2-7 de la revue de littérature reprenant la matrice pedigree

Onglet *Papers* : Liste des données publiées et utilisées dans le cadre de ce projet

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
Kim,S. and Overcash, M.	2003	Paper	Energy in chemical manufacturing process: gate-to-gate information for life cycle assessment			Journal of Chemical Technology and Biotechnology			78	995-1005	
IVF - Industrial Research and Development Corporation	2007	Report	Lot 3 Personal Computers (desktops and laptops) and Computer Monitors	IVF Report 07004	1404-191X		Sweden	European Commission DG TREN			http://www.eup-network.de/fileadmin/user_upload/Produktgruppen/Lots/Final_Documents/EuP_Lot3_PC_FinalReport.pdf
Milota,M.R.; West,C.D.; Hartley,I.D.	2005	Paper	Gate-to-Gate Life Cycle Inventory of Softwood Lumber Production			Wood and Fiber Science			37	47-57	
Wilson, J.B.	2010	Paper	Life-Cycle Inventory of formaldehyde-based resins used in wood composites in terms of resources, emissions, energy and carbon			Wood and Fiber Science			42	125-143	
Wagner, F.G.; Puetmann, M.E.; Johnson, L.R.	2009	Report	Life cycle inventory of inland Northwest softwood lumber manufacturing	CORRIM: Phase II Final report - Module B			United States	CORRIM			http://www.corrim.org/pubs/reports/2010/phase2/module_b.pdf
Wilson, J.B.	2008	Report	Particleboard: a life-cycle inventory of	CORRIM: Phase II Final report			United States	CORRIM			http://www.corrim.org/pubs/reports/2010/phase2/

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			manufacturing panels from resource through product	- Module F							module_f.pdf
Hubbard, S. S. ; Bowe, S.A.	2008	Report	Life-cycle inventory of solid strip hardwood flooring in the Eastern United States	CORRIM: Phase II Final report - Module E			United States	CORRIM			http://www.corrim.org/pubs/reports/2010/phase2/module_e.pdf
Athena Institute	2009	Report	A cradle to gate life cycle assessment of Canadian medium density fiberboard	Final Report			Canada	Athena Sustainable Materials Institute			
Gac, A. ; Deltour, L. ; Cariolle, M. ; Dollé, J.B. ; Espagnol, S. ; Flénet, F. ; Guinguand, N. ; Lagadec, S. ; Le Gall, A. ; Lellahi, A. ; Malaval, C. ; Ponchant, P. ; Tailleur, A.	2010	Report	GES'TIM Guide méthodologique pour l'estimation des impacts des activités agricoles sur l'effet de serre	V 1.2			France	Institut de l'élevage - IFIP - ITAVI - ARVALIS - CETIOM - ITB			http://idele.fr/contact/publication/idelesolr/recommends/gestim-guide-methodologique-pour-lestimation-des-impacts-des-activites-agricoles-sur-leffet-de/print.html
Schorcht, F. ; Kourti, I. ; Scalet, B.M. ; Roudier, S. ; Delgado Sancho, L.	2013	Report	Best available techniques (BAT) reference document for the production of cement, lime and magnesium oxide	Report EUR 26129 EN	1831-9424		European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/CLM_Published_def.pdf

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
European Commission	2007	Report	Reference document on the Best Available Techniques in the Ceramic Manufacturing Industry	CER			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/cer_bref_0807.pdf
European Commission	2006	Report	Integrated Pollution Prevention and Control. Reference Document on Best Available Techniques in the Food, Drink and Milk Industries	FDM			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/fdm_bref_0806.pdf
European Commission	2003	Report	Integrated Pollution Prevention and Control Reference Document on Best Available Techniques for Intensive Rearing of Poultry and Pigs	ILF			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/irpp_bref_0703.pdf
Remus,R. ; Aguando Monsonet, M.A. ; Roudier,S. ; Delgado Sabcho,L.	2013	Report	Bests Available Techniques (BAT) Reference Document for Iron and Steel Production	Report EUR 25521 EN	1831-9424		European Union	Joint Research Center			https://ec.europa.eu/jrc/sites/default/files/lfna25521enn_0.pdf
European Commission	2007	Report	Reference Document on Best Available Techniques for	LVIC-S			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/lvic-

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			the Manufacture of Large Volume Inorganic Chemicals - Solids and Others industry								s_bref_0907.pdf
European Commission	2007	Report	Reference Document on Best Available Techniques for the Manufacture of Large Volume Inorganic Chemicals - Ammonia, Acids and Fertilisers	LVIC-AAF			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/lvic_aaf.pdf
European Commission	2007	Report	Reference Document on Best Available Techniques in the Production of Polymers	POL			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/pol_bref_0807.pdf
European Commission	2001	Report	Reference Document on Best Available Techniques in the Pulp and Paper Industry	PPM			European Union	Joint Research Center			eippcb.jrc.ec.europa.eu/reference/BREF/ppm_bref_1201.pdf
European Commission	2005	Report	Reference Document on Best Available Techniques in the Smitheries and Foundries Industry	SF			European Union	Joint Research Center			eippcb.jrc.ec.europa.eu/reference/BREF/sf_bref_0505.pdf
European Commission	2003	Report	Reference Document of Best Available Techniques for	TXT			European Union	Joint Research Center			eippcb.jrc.ec.europa.eu/reference/BREF/txt_bref_0703.pdf

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			the Textile Industry								
Özilgen,M. ; Sorgüven,E.	2011	Paper	Energy and exergy utilization, and carbon dioxide emission in vegetable oil production		0360-5442	Energy			36 (10)	5954-5967	http://www.sciencedirect.com/science/article/pii/S0360544211005500
Elsam Engineering A/S	2004	Report	Life Cycle Assessment of offshore and onshore sited wind farms	T012063			Denmark	Elsam Engineering A/S			http://www.aper.org/manager/docnum/doc/doc1252_LCA_V80_2004_uk%5B1%5D.fiche%2042.pdf
Vestas	2006	Report	Life cycle assessment of electricity produced from onshore sited wind power plants based on Vestas V82-1.65 MW turbines				Denmark	Vestas Wind Systems A/S Alsvej 21 8900 Randers			http://www.vestas.com/files/filer/en/sustainability/lca/lcav82165mwonshore.pdf
Vestas	2006	Report	Life cycle assessment of offshore and onshore sited wind power plants based on Vestas V90-3.0MW turbines				Denmark	Vestas Wind Systems A/S Alsvej 21 8900 Randers			http://www.vestas.com/Files/Filer/EN/Sustainability/LCA/LCAV90_juni_2006.pdf
Teehan,P. ; Kandikar,M.	2012	Paper	Sources of Variation in Life Cycle Assessments of Desktop Computers		1530-9290	Journal of Industrial Ecology			16	S182-S194	

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
Healy,M.L. ; Dahlben,L.J ; Isaacs,J.A.	2008	Paper	Environmental Assessment of Single-Walled Carbon Nanotube Processes		1530-9290	Journal of Industrial Ecology			12 (3)	376-393	
Grubb,G.F. ; Bakshi,B.R.	2011	Paper	Life Cycle of Titanium Dioxide Nanoparticle Production		1530-9290	Journal of Industrial Ecology			15 (1)	81-95	
Iriarte,A. ; Rieradevall, J. ; Gabarrell, X.	2010	Paper	Life cycle assessment of sunflower and rapeseed as energy crops under Chilean conditions		0959-6526	Journal of Cleaner Production			18(4)	336-345	
Goglio,P. ; Bonari,E. ; Mazzoncini, M.	2012	Paper	LCA of cropping systems with different external input levels for energetic purposes		0961-9534	Biomass and Bioenergy			42(0)	33-42	
Chiaramonti ,D. ; Recchia,L.	2010	Paper	Is life cycle assessment (LCA) a suitable method for quantitative CO2 saving estimations? The impact of field input on the LCA results for a pure vegetable oil chain		0961-9534	Biomass and Bioenergy			34(5)	787-797	
Franklin Associates	2010	Report	Life cycle inventory of 100% postconsumer HDPE and PET	Revised final report CLIENTS/ACC/KC102			United States	Franklin Associates, a division of ERG			http://www.napco.com/pdf/Final Report_LCI_Pos tconsumer_PET

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			recycled resin from postconsumer containers and packaging	305							andHDPE.pdf
Sheehan,J. ; Camobresco,V. ; Duffield,J. ; Graboski,M. ; Shapouri,H.	1998	Report	Life Cycle Inventory of Biodiesel and Petroleum Diesel for Use in an Urban Bus	NREL/SR-580-24089 UC Category 1503			United States	National Renewable Energy Laboratory - US Departemen t of Agriculture			http://www.nrel.gov/docs/legosti/fy98/24089.pdf
Remy,C.	2010	Thesis	Life cycle assessment of conventional and source-separation systems for urban wastewater management				Berlin	Technischen Universität Berlin - Fakultät III - Prozesswissenschaften			
Franklin Associates	2004	Report	Life cycle inventory of packaging options for shipment of retail mail-order soft goods	CLIENTSVK C041682.doc			United States	Franklin Associates, a division of ERG			http://www.deq.state.or.us/lq/pubs/docs/sw/packaging/LifeCycleInventory.pdf
FEFCO	2009	Report	European Database for Corrugated Board Life Cycle Studies					FEFCO - Cepi ContainerBoard			http://www.fefco.org/sites/default/files/documents/2009_LCA_report_Fefco.pdf
Espinosa,N. ; Garcia-Valverde,R. ; Krebs,F.C.	2011	Paper	Life-cycle analysis of product integrated polymer solar cells		1754-5692	Energy & Environmental Science			4(5)	1547-1557	
Blonk,H. ; Ponsioen,T.	2011	Report	The Agri-footprint method -				The Netherlands	Blonk Milieu Advies BV			

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
; Kool,A. ; Marinussen, M.			Methodological LCA framework, assumptions and applied data - Annex				s				
Bureau of transportati on statistics - Resaerch and Innovation Technology Administrati on	2011	Report	National Transportation statistics - 2011				United States	U.S. Department of Transportati on.			http://www.bts.gov/publications/national_transportation_statistics/
Ibáñez- Forés,V. ; Bovea,M.D. ; Simò,A.	2011	Paper	Life cycle assessment of ceramic tiles. Environmental and statistical analysis		0948- 3349	The Internatio nal Journal of Life Cycle Assessme nt			16(9)	916-928	
Fthenakis,V .; Kim,H.C.	2009	Paper	Land use and electricity generation: A life- cycle analysis		1364- 0321	Renewabl e and Sustainabl e Energy Reviews			13(6-7)	1465- 1474	
Alverbro,K. ; Björklund,A. ; Finnveden, G. ; Hochschorn er,E. ; Häggvall,J.	2009	Paper	A life cycle assessment of destruction of ammunition		0304- 3894	Journal of Hazardou s Materials			170(2- 3)	1101- 1109	
Koehler,A. ; Wildblotz,C.	2009	Paper - Supporting information	Comparing the Environmental Footprints of Home-Care and Personal-Hygiene		0013- 936X	Environm ental Science & Technolog y			43(22)	8643- 8651	

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			Products: The Relevance of Different Life-Cycle Phases								
International Aluminium Institute	2003	Report	Life Cycle Assessment of Aluminium: Inventory Data for the Worldwide Primary Aluminium Industry					International Aluminium Institute			http://transport.world-aluminium.org/uploads/media/1274452849Global_LCI_Report.pdf
Venta,G.J.	2001	Report	Cradle-to-gate life cycle inventory for exterior stucco finishes				Canada	The Athena Sustainable Materials Institute			http://calculatelca.com/wp-content/themes/athena/images/LCA%20Reports/Exterior_Stucco.pdf
Meil,J.K.	1998	Report	A Life Cycle Analysis of Solid Wood and Steel Cladding				Canada	The Athena Sustainable Materials Institute			http://calculatelca.com/wp-content/themes/athena/images/LCA%20Reports/Solid_Wood_And_Steel_Cladding.pdf
Norgate,T. ; Haque,N.	2012	Paper	Using life cycle assessment to evaluate some environmental impacts of gold production		0959-6526	Journal of Cleaner Production			29-30(0)	53-63	
Lükewill,A. ; Bertok,I ; Amann,M. ; Cofala,J. ; Gyarfas,F. ; Heyes,C. ;	2001	Report	A Framework to Estimate the Potential and Costs for the Control of Fine Particulate	Interim Report - IR 01 023			Austria	International Institute for Applied System Analysis			http://www.iiasa.ac.at/~rains/reports/IR-01-023.pdf

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
Karvosenoj a.N ; Klimont,Z. ; Schöpp,W.			Emissions in Europe								
Department for transport - United Kingdom	2009	Database	Van CO ₂ and Fuel Consumption database				United Kingdom	Department for transport			http://vanfueldat.a.dft.gov.uk/Default.aspx
Hubbard,S. S. ; Bowe,S.A.	2010	Paper	A Gate-to-Gate Life Cycle Inventory of Solid Hardwood Flooring in the eastern US			Wood and Fiber Science			42	79-89	
ISPRA	2013	Report	Italian Greenhouse Gas Inventory 1990-2011. Nationale Inventory report 2013	177/2013	978-88-448-0609-5		Italy	Institute for Environmental Protection and Research			http://www.isprambiente.gov.it/en/publications/reports/italian-greenhouse-gas-inventory-1990-2011.-national-inventory-report-2013
Garrett,P. ; Ronde,K.	2011	Report	Life Cycle Assessment of Electricity Production from a V80-2.0MW Gridstreamer Wind Plant				Denmark	Vestas Wind Systems A/S Alsvej 21 8900 Randers			http://www.vestas.com/files%2Ffiler%2Fen%2Fsustainability%2Flica%2Flica_v80-2mw_version1.pdf
Garrett,P. ; Ronde,K.	2011	Report	Life Cycle Assessment of Electricity Production from a V100-1.8MW Gridstreamer				Denmark	Vestas Wind Systems A/S Alsvej 21 8900 Randers			http://www.vestas.com/~media/vestas/about/sustainability/pdfs/lica_v100-1_8mw_onshore

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			Wind Plant								_gridstreamer_version1_2011.pdf
Natural Ressources Canada	2010	Database	GHGenius: A model for lifecycle assessment of transportation fuels	Version 3.19			Canada	Natural Ressources Canada			http://www.ghgenius.ca/
Bretscher,D .	2010	Report	Agricultural CH4 and N2O emissions in Switzerland	QA/QC			Switzerland	Federal Department of Economic Affairs - Agroscope Reckenholz Tänikon Research Station			http://bit.ly/1Vbj0T
Trozzi,C. ; Rentz,O. ; Oertel,D. ; Woodfield, M. ; Stear,R.	2010	Report	Energy industries - Combustion in energy and transformation industries	EMEP/EEA emission inventory guidebook 2009, updated June 2010			European Union	Europe Environment Agency			http://www.eea.europa.eu/publications/emep-eea-emission-inventory-guidebook-2009
European Commission	2006	Report	Reference Document on Best Available Techniques for Large Combustion Plant	LCP			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/lcp_bref_0706.pdf
European Commission	2003	Report	Reference Document on Best Available Techniques in the Large Volume Organic Chemical	LVOC			European Union	Joint Research Center			http://www.elaw.org/system/files/LargeVolumeOrganicChemicalIndustry_0.pdf

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
			Industry								
Scalet,B.M. ; Garcia Munoz,M. ; Sossa,A.Q. ; Roudier,S. ; Delgado Sancho,L.	2013	Report	Best Available Techniques Reference Document for the Manufacture of Glass	EUR 25786 EN	1831-9424		European Union	Joint Research Center			
European Commission	2003	Report	Reference Document on Best Available Techniques for Mineral Oil and Gas Refineries	REF			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/reference_bref_0203.pdf
European Commission	2001	Report	Reference Document on Best Available Techniques in the Non Ferrous Metals Industries	NFM			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/nfm_bref_1201.pdf
European Commission	2006	Report	Reference Document on Best Available Techniques for the Waste Treatment Industries	WT			European Union	Joint Research Center			http://eippcb.jrc.ec.europa.eu/reference/BREF/wt_bref_0806.pdf
D'Souza,N. ; Gbegbaje-Das,E. ; Shonfield,P.	2011	Report	Life Cycle Assessment of Electricity Production from a V112 Turbine Wind Plant				Denmark	PE North West Europe ApS			http://www.vestas.com/~media/vestas/about/sustainability/pdfs/lca_v112_study_report_2011.pdf

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
Environment Canada	2005	Excel Database	Combustion du charbon anthraciteux (chaudières à alimentation automatique)				Canada	Environment Canada			https://www.ec.gc.ca/inrpnpri/9C8F5570-19B6-4E37-97DB-290D4CE28153/antracite_coal_combustion_f_04_02_2009.xls
International Energy Agency	2012	Report	CO2 emissions from fuel combustion - Highlights				France	OECD/IEA			
European Commission	2003	Report	Document de référence sur les meilleures techniques disponibles - Tannage des cuirs et peaux	TAN			European Union	Joint Research Center			https://www.yumpu.com/fr/document/view/6892162/bref-tannage-des-peaux-document-integral-ineris
US Environmental Protection Agency	1995	Report	Compilation of air pollutant emission factors	AP-42 Fifth edition - Appendix B			United States	USEPA - Office of air quality planning and standards - Office of air and radiation			http://www.epa.gov/ttnchie1/ap42/appendix/appb-1.pdf
Berdowski, J. ; Visschedijk, A. ; Creemers, E. ; Pulles, T.	1995	Excel Database	CEPMEIP - Emission factors for particulate matter - Reference document				The Netherlands	TNO-MEP			http://www.air.sk/tno/cepmeip/
Natural Resources Canada	2008	Excel Database	National Energy Use Database				Canada	Natural Resources Canada			http://oee.nrcan.gc.ca/corporate/statistics/neud/d

Authors	Year	Type of publication	Title	Report Number	ISSN/ ISBN	Journal	Place published	Publisher	Volume	Pages	URL
											pa/home.cfm?att r=0

Onglet *DataClassification* : Comment les données ont été classées dans les différents couples « indicateur pedigree / score pedigree »

Pedigree indicator	Score 1	Score 2	Score 3	Score 4	Score 5
Reliability	Data based on calculations or measurements	Score 1 + Data based on verified data sources (statistics, scientific paper, specific database)	Score 2 + Data from generic database or default value	Score 3 + Data based on model, estimation or expert judgment	Score 4 + Unverified data
Temporal correlation	Data from 2010 to 2013	Score 1 + data from 2007 to 2009	Score 2 + data from 2001 to 2006	Score 3 + data from 2001 to 1998	Score 4 + Data from all other years
Geographical correlation	Data from United States	Score 1 + data from Canada and Mexico	Score 2 + data from Latin America	Score 3 + data from Eurasia and Oceania	Score 4 + data from the rest of the world

Pour le cas du critère de corrélation technologique

Industrial sector	Reference product	Score 1	Score 2	Score 3	Score 4	Score 5
Agriculture	<i>Corn production</i>	Data from corn production	Score 1 + data from cereal production (NAICS: 111150)	Score 2 + data from crop production (NAICS 1115)	Score 3 + data from all vegetables production (NAICS 111)	Score 4 + data from the whole agricultural sector (NAICS 11)
	<i>Dairy cattle</i>	Data from dairy cattle from one data source	Score 1 + data from all dairy cattle (NAICS 112120)	Score 2 + data from cattle ranching (NAICS 1121)	Score 3 + Data from all animal production (NAICS 112)	<i>Not applicable</i>
Chemical products manufacturing	<i>Titanium dioxide production by chloride process</i>	Data from titanium production by chloride process	Score 1 + data from titanium production	Score 2 + data from inorganic chemicals production (NAICS 325180)	Score 3 + data from all basic chemical manufacturing (NAICS 3251)	Score 4 + data from chemical manufacturing (NAICS 325)
Combustion	<i>Hard coal combustion in the iron and steel sector</i>	Data from coal combustion in the iron and steel sector	Score 1 + data from coal combustion in other industrial sectors	Score 2 + data from coal combustion	Score 3 + data from fossil fuel combustion	Score 4 + data from all type of fuels combustion
Electricity production	<i>Onshore wind electricity production</i>	Data from onshore wind electricity production for one data source	Score 1 + data from onshore wind electricity production	Score 2 + wind electricity production	Score 3 + data from all renewable electricity production	Score 5 + data from all type of electricity sources
	<i>Electricity production by coal</i>	Data from electricity production by coal for 1 data source	Score 1 + data from electricity production by coal	Score 2 + data from electricity production by fossil fuel (NAICS	Score 3 + data from electricity production by non- renewables	Score 5 + data from all type of electricity sources

Industrial sector	Reference product	Score 1	Score 2	Score 3	Score 4	Score 5
				221112)	resource	
Metal manufacturing	<i>Cast iron production</i>	Data from cast iron production from one data source	Score 1 + data from iron production	Score 2 + data from iron and steel production	Score 3 + data from ferrous metal production	Score 4 + data from primary metal production (NAICS 331)
Mining	<i>Coal mining</i>	Data from coal mining for one data source	Score 1 + data from coal mining	Score 2 + data related to coal mining (NAICS 2121)	Score 3 + data from clay and gold mining	Score 4 + data from all mining activities (NAICS 21)
Transport	<i>Passengers cars</i>	Data from passenger cars for one type fuel	Score 2 + data from passenger cars for different type of fuels	Score 3 + data from all type of passenger road transportation	Score 4 + data from road transportation	Score 5 + data from the transportation sector
Transport operations	<i>Passenger cars fuel operation</i>	Data from passenger cars fuel operation	Score 1 + data from light duty vehicles operation	Score 2 + data from motorcycle et 4x4 operation	Score 3 + data from light duty trucks operation	Score 4 + data from all road transportation operation
Waste	<i>Incineration of municipal solid waste</i>	Data from municipal solid waste incineration	Score 1 + data from waste incineration	Score 2 + data from open burning waste management	Score 3 + data from all waste treatment	<i>Not applicable</i>

Onglet *AUF_1* Données utilisées pour obtenir la distribution a priori des facteurs d'incertitude additionnelle

Paper or report citation	Data specific ID	Reliability, Score 2	Reliability, Score 3	Reliability, Score 4	Reliability, Score 5
Weidema and Wesnaes (1996)	Weidema et Wenaes 1996 (min)	1.06	1.40	1.64	5.10
	Weidema et Wenaes 1996 (max)	1.22	1.64		
Frischknecht et al. (2005)	DQG_ecoinvent_v0.14	1.05	1.10	1.20	1.50
Ciroth et al. (2012)	Ciroth_Gemis	1.54	1.61	1.69	43.55
	Ciroth_PRTR		1.02	1.37	

Paper or report citation	Data specific ID	Completeness, Score 2	Completeness, Score 3	Completeness, Score 4	Completeness, Score 5
Weidema and Wesnaes (1996)	Weidema et Wenaes 1996 (min)	1.00			1.64
	Weidema et Wenaes 1996 (max)	1.22			
Frischknecht et al. (2005)	DQG_ecoinvent_v0.14	1.02	1.05	1.10	1.20
Ciroth et al. (2012)	Ciroth_Gemis				
	Ciroth_PRTR				
	Ciroth_Yoghurts	1.03	1.04	1.08	

Paper or report citation	Data specific ID	Temporal correlation, Score 2	Temporal correlation, Score 3	Temporal correlation, Score 4	Temporal correlation, Score 5
Weidema and Wesnaes (1996)	Weidema et Wenaes 1996 (min)	1.22	1.49	1.80	2.16
	Weidema et Wenaes 1996 (max)				
Frischknecht et al. (2005)	DQG_ecoinvent_v0.14	1.03	1.10	1.20	1.50
Ciroth et al. (2012)	Ciroth_NATransportationStatistics	1.00	1.01	1.02	1.02
	Ciroth_NATransportationStatistics_Air	1.01	1.05	1.07	1.08
	Ciroth_NATransportationStatistics_General	1.01	1.05	1.08	1.08
	Ciroth_NATransportationStatistics_Marine	1.00	1.00	1.01	1.01
	Ciroth_NATransportationStatistics_Others	1.02	1.03	1.05	1.04
	Ciroth_NATransportationStatistics_Rail	1.01	1.00	1.01	1.01

Paper or report citation	Data specific ID	Temporal correlation, Score 2	Temporal correlation, Score 3	Temporal correlation, Score 4	Temporal correlation, Score 5
	Ciroth_NATransportationStatistics_Road	1.02	1.04	1.10	1.12
	Ciroth_Tremod	1.03	1.10	1.19	1.29
	Ciroth_Tremod_Benzene	1.06	1.09	1.09	1.08
	Ciroth_Tremod_Methane	1.12	1.22	1.28	1.30
	Ciroth_Tremod_CarbonMonoxide	1.03	1.01	1.02	1.04
	Ciroth_Tremod_CO2(rep)	1.01	1.02	1.01	1.01
	Ciroth_Tremod_CO2tot	1.01	1.01	1.01	1.01
	Ciroth_Tremod_FC	1.01	1.01	1.01	1.01
	Ciroth_Tremod_HC	1.10	1.18	1.22	1.24
	Ciroth_Tremod_N2O	1.09	1.11	1.12	1.12
	Ciroth_Tremod_NH3	1.10	1.17	1.15	1.09

Paper or report citation	Data specific ID	Temporal correlation, Score 2	Temporal correlation, Score 3	Temporal correlation, Score 4	Temporal correlation, Score 5
	Ciroth_Tremod_NMHC	1.06	1.10	1.12	1.13
	Ciroth_Tremod_NO2	1.03	1.05	1.06	1.06
	Ciroth_Tremod_NOx	1.02	1.02	1.02	1.02
	Ciroth_Tremod_Pb	1.00	1.02	1.01	1.75
	Ciroth_Tremod_PM	1.08	1.15	1.25	1.32
	Ciroth_Tremod_PN	1.06	1.07	1.12	1.15
	Ciroth_Tremod_SO2	1.19	1.73	2.88	3.86

Paper or report citation	Data specific ID	Geographical correlation, Score 2	Geographical correlation, Score 3	Geographical correlation, Score 4	Geographical correlation, Score 5
Weidema and Wesnaes (1996)	Weidema et Wenaes 1996 (min)	1.11	1.22	2.57	2.57
	Weidema et Wenaes 1996 (max)	1.40	1.64		
Frischknecht et al. (2005)	DQG_ecoinvent_v0.14	1.01	1.02	1.05	1.10
Ciroth et al. (2012)	Ciroth_PRTR	1.04	1.08	1.11	
	Ciroth_NATransportationStatistics	1.16	1.48		
	Ciroth_Tremod	1.02			
	Ciroth_Greet	1.00		1.03	

Paper or report citation	Data specific ID	Further technological correlation, Score 2	Further technological correlation, Score 3	Further technological correlation, Score 4	Further technological correlation, Score 5
Weidema and Wesnaes (1996)	Weidema et Wenaes 1996 (min)	1.37	1.64	1.90	2.57
	Weidema et Wenaes 1996 (max)				
Frischknecht et al. (2005)	DQG_ecoinvent_v0.14	1.05	1.20	1.50	2.00
Ciroth et al. (2012)	Ciroth_Tremod	1.18	1.65	2.08	2.80
	Ciroth_Greet	1.00	1.02	1.07	

ONGLET AUF_2 Facteurs d'incertitude additionnelle obtenus selon les différentes hypothèses faites sur la distribution à priori

1st assumption: σ_{1UF} is defined based on the 95th confidence interval (see Figure 6-2)

2nd assumption: σ_{2UF} is defined based on the 66th confidence interval (see Figure 6-2)

	FIRST ASSUMPTION				SECOND ASSUMPTION			
	<i>UF2</i>	<i>UF3</i>	<i>UF4</i>	<i>UF5</i>	<i>UF2</i>	<i>UF3</i>	<i>UF4</i>	<i>UF5</i>
RELIABILITY								
Generic	1.01	1.21	1.25	2.36	1.01	1.27	1.30	2.55
Agriculture	1.00	1.05	1.07	1.60	1.00	1.05	1.05	1.64
Combustion	1.06	1.13	1.18	1.69	1.06	1.14	1.17	1.98
Manufacturing	1.00	1.20	1.25	2.60	1.00	1.32	1.32	3.09
Metal Manufacturing	1.00	1.00	1.21	1.51	1.00	1.00	1.23	1.51
COMPLETENESS								
Generic	1.02	1.05	1.10	1.63	1.02	1.05	1.11	2.23
TEMPORAL CORRELATION								
Generic	1.09	1.55	2.22	2.49	1.17	1.92	2.49	2.63
Agriculture	1.03	1.13	1.80	4.07	1.05	1.23	2.81	3.44
Combustion	1.08	1.27	1.72	1.75	1.16	1.49	1.98	2.24
Utilities	1.00	1.22	1.87	2.52	1.00	1.49	2.94	3.53
Manufacturing	1.05	1.37	2.12	2.47	1.09	1.96	2.97	3.33
Metal Manufacturing	1.07	1.22	1.32	1.49	1.11	1.30	1.34	1.49
Transportation	1.16	1.26	1.26	1.15	1.19	1.29	1.26	1.15
GEOGRAPHICAL CORRELATION								
Generic	1.14	1.23	2.36	2.30	1.25	1.34	3.24	3.19
Agriculture	1.00	1.10	1.57	1.55	1.00	1.22	2.23	2.21
Combustion	1.08	1.11	1.66	1.65	1.13	1.15	2.15	2.13
Manufacturing	1.09	1.16	1.98	1.92	1.25	1.38	3.76	3.59
FURTHER TECHNOLOGICAL CORRELATION								
Generic	1.19	1.52	1.95	2.23	1.34	1.88	2.30	2.60
Agriculture	1.00	1.32	1.59	2.03	1.00	1.58	1.77	2.10
Combustion	1.04	1.04	1.50	1.89	1.04	1.03	1.50	1.73
Utilities	1.00	1.29	1.62	2.01	1.00	1.43	1.78	2.01
Chemical manufacturing	1.16	1.44	2.04	2.21	1.43	2.02	3.02	2.78

ONGLET AUF_3 Données utilisées pour définir les fonctions de vraisemblance des facteurs d'incertitude additionnelle

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Further technological correlation	Agriculture	Corn production	Product exchange		Nitrogen fertilizer	kg/m2	1.00	1.00	6.70	6.70	6.70
Further technological correlation	Agriculture	Corn production	Product exchange		Seed	kg/m2	1.00	1.00	8.47	8.47	8.47
Further technological correlation	Agriculture	Corn production	Product exchange		Potassium oxide	kg/m2	1.00	1.00	1.44	1.44	1.44
Further technological correlation	Agriculture	Corn production	Product exchange		Herbicides	kg/m2	1.00	1.00	2.27	0.44	2.27
Further technological correlation	Agriculture	Corn production	Product exchange		Phosphorous pentoxide	kg/m2	1.00	1.00	1.03	1.03	1.03
Further technological correlation	Agriculture	Corn production	Product exchange		Water	kg/m2	1.00	1.00	1.04	1.04	1.04
Further technological correlation	Agriculture	Dairy cattle	Emission	Air	Methane	kg/animal/yr	1.00	1.00	1.75	3.85	
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	unspecified PM	kg/kg	1.00	4.44	4.44	4.44	74.81
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	Chlorine	kg/kg	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Chemical production	Titanium dioxide production	Waste		Waste	kg/kg	1.00	1.00	1.07	1.07	3.26
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	Carbon dioxide	kg/kg	1.00	1.47	1.55	6.11	6.95
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Chlorinated organic compounds	kg/kg	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Lead	kg/kg	1.00	1.53	27.52	27.52	21.47
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Suspended solids, unspecified	kg/kg	1.00	9.74	79.27	79.27	125.56
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Vanadium	kg/kg	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Chemical production	Titanium dioxide production	Product exchange		Energy	J/kg	1.00	1.38	2.70	2.70	15.03
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Nickel	kg/kg	1.00	1.33	1.33	0.75	2.31
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Zinc	kg/kg	1.00	19.12	19.12	19.12	38.82
Further technological correlation	Chemical	Titanium dioxide production	Emission	Air	Nitrogen oxide	kg/kg	1.00	1.40	1.11	1.68	1.91

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
correlation	production										
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	Sulfur oxide	kg/kg	1.00	27.03	20.22	19.35	25.78
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Chromium	kg/kg	1.00	17.52	17.52	17.52	25.41
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Hydrochloric acid	kg/kg	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Iron	kg/kg	1.00	47.27	47.27	47.27	47.27
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Manganese	kg/kg	1.00	10.13	10.13	10.13	10.13
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Mercury	kg/kg	1.00	1.56	3.62	3.62	13.46
Further technological correlation	Chemical production	Titanium dioxide production	Product exchange		Water	m3/kg	1.00	1.00	1.00	9.08	25.92
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	Carbon monoxide	kg/kg	1.00	63.61	96.12	60.85	275.64
Further technological correlation	Chemical production	Titanium dioxide production	Emission	Air	Carbonyl sulphide	kg/kg	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Cadmium	kg/kg	1.00	1.29	1.29	1.29	2.53
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Copper	kg/kg	1.00	2.60	2.60	2.60	6.77
Further technological correlation	Chemical production	Titanium dioxide production	Emission	water	Titanium	kg/kg	1.00	6.75	6.75	6.75	6.75
Further technological correlation	Combustion	Hard coal combustion in iron and steel industry	Emission	air	PM10	kg/kg	1.00	1.03	1.01	1.92	1.92
Further technological correlation	Combustion	Hard coal combustion in iron and steel industry	Emission	air	Suspended solids, unspecified	kg/kg	1.00	1.04	1.00	1.57	1.57
Further technological correlation	Combustion	Hard coal combustion in iron and steel industry	Emission	air	PM2.5	kg/kg	1.00	1.05	1.09	1.06	1.08
Further technological correlation	Electricity generation	By coal	Emission	air	Carbon dioxide	kg/J	1.00	1.00	1.08	1.35	2.00
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Nitrogen fertilizer	kg/J	1.00	1.00	1.68	0.59	1.68
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Carbon dioxide	kg/J	1.00	1.00	4314.04	6519.52	3.04
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Nitrogen oxide	kg/J	1.00	1.00	2.72	6.37	7.33
Further technological correlation	Electricity	By onshore wind production	Emission	air	NMVOC	kg/J	1.00	1.00	1.11	1.02	1.58

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
correlation	generation										
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	PM2.5	kg/J	1.00	1.00	1.06	13.91	12.02
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	unspecified PM	kg/J	1.00	1.00	1.09	8.03	7.29
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Chloride	kg/J	1.00	1.00	1.06	0.94	1.06
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Iron	kg/J	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Sodium component	kg/J	1.00	1.00	1.68	1.68	1.68
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Suspended solids, unspecified	kg/J	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Electricity generation	By onshore wind production	Resource		Aluminium	kg/J	1.00	1.00	1.16	1.16	1.16
Further technological correlation	Electricity generation	By onshore wind production	Resource		Coal	kg/J	1.00	1.00	1.09	0.92	1.09
Further technological correlation	Electricity generation	By onshore wind production	Resource		Iron	kg/J	1.00	1.00	1.65	0.61	1.65
Further technological correlation	Electricity generation	By onshore wind production	Resource		Natural Gas	kg/J	1.00	1.00	1.06	1.06	1.06
Further technological correlation	Electricity generation	By onshore wind production	Resource		Oil	kg/J	1.00	1.00	1.05	0.95	1.05
Further technological correlation	Electricity generation	By onshore wind production	Resource		Sodium component	kg/J	1.00	1.00	1.37	1.37	1.37
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Carbon monoxide	kg/J	1.00	1.00	2.82	0.35	2.33
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Hydrogen	kg/J	1.00	1.00	1.31	1.31	1.31
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Steam	kg/J	1.00	1.00	1.23	0.81	1.23
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Oxygen	kg/J	1.00	1.00	1.14	1.14	1.14
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Sulfur oxide	kg/J	1.00	1.00	1.04	12.93	52.17
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Methane	kg/J	1.00	1.00	1.57	11.60	9.17
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	COD	kg/J	1.00	1.00	2.51	2.51	2.51
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Sodium	kg/J	1.00	1.00	1.15	0.87	1.15

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
correlation	generation										
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Dinitrogen	kg/J	1.00	1.00	1.03	1.60	5.18
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	Dust	kg/J	1.00	1.00	1.07	1.07	1.07
Further technological correlation	Electricity generation	By onshore wind production	Emission	air	VOC	kg/J	1.00	1.00	3.47	3.47	3.47
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Calcium	kg/J	1.00	1.00	1.01	0.99	1.01
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Fluoride	kg/J	1.00	1.00	1.41	1.41	1.41
Further technological correlation	Electricity generation	By onshore wind production	Emission	water	Sulfate	kg/J	1.00	1.00	1.12	0.90	1.12
Further technological correlation	Electricity generation	By onshore wind production	Product exchange		Electricity	J/J	1.00	1.00	1.28	1.28	1.28
Further technological correlation	Electricity generation	By onshore wind production	Resource		Lime and Limestone	kg/J	1.00	1.00	1.19	0.84	1.19
Further technological correlation	Electricity generation	By onshore wind production	Resource		Water	kg/J	1.00	1.00	1.03	0.97	1.03
Further technological correlation	Metal Manufacturing	Cast iron	Emission	air	Benzo-fluoranthene	kg/m3	1.00	1.00	1.00	1.00	1.00
Further technological correlation	Metal Manufacturing	Cast iron	Emission	air	Dioxin	kg/m3	1.00	1.00	1.13	1.48	1.48
Further technological correlation	Mining	Coal mining	Emission	air	Methane	m3/kg	1.00	1.39	1.64	1.64	1.64
Further technological correlation	Transport	Passenger cars	Emission	Air	Suspended solids, unspecified	kg/m	1.00	1.11	1.11	1.11	1.05
Further technological correlation	Vehicle operations	Passenger cars fuel operation	Product exchange		Operation	m/m3	1.00	1.00	2.01	2.01	3.14
Geographical correlation	Combustion		Emission	air	NM VOC	kg/J	1.00	1.00	1.00	1.59	1.59
Geographical correlation	Combustion		Emission	air	Carbon monoxide	kg/m3	1.00	1.00	1.00	21.21	21.21
Geographical correlation	Combustion		Emission	air	Sulfur oxide	kg/kg	1.00	1.00	1.00	5.51	5.51
Geographical correlation	Combustion		Emission	air	Nitrogen oxide	kg/kg	1.00	1.00	1.00	1.61	1.61
Geographical	Combustion		Emission	air	Carbon	kg/kg	1.00	1.00	1.00	2.71	2.71

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
correlation					monoxide						
Geographical correlation	Combustion		Emission	air	unspecified PM	kg/kg	1.00	1.00	1.00	1.40	1.40
Geographical correlation	Combustion		Emission	air	Nickel	kg/kg	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Combustion		Emission	air	Methane	kg/J	1.00	1.17	1.17	1.05	1.05
Geographical correlation	Combustion		Emission	air	Copper	kg/J	1.00	1.00	1.00	1.08	1.08
Geographical correlation	Combustion		Emission	air	Sulfur oxide	kg/J	1.00	2.27	2.27	2.65	2.65
Geographical correlation	Combustion		Emission	air	Zinc	kg/J	1.00	1.00	1.00	1.30	1.30
Geographical correlation	Combustion		Emission	air	Nitrogen oxide	kg/J	1.00	1.17	1.17	1.23	1.23
Geographical correlation	Combustion		Emission	air	Nitrogen oxide	kg/m3	1.00	1.00	1.00	108.30	108.30
Geographical correlation	Combustion		Emission	air	PM2.5	kg/J	1.00	1.08	1.08	1.07	1.07
Geographical correlation	Combustion		Emission	air	Carbon monoxide	kg/J	1.00	1.31	1.31	1.71	1.71
Geographical correlation	Combustion		Emission	air	Sulfur oxide	kg/m3	1.00	1.00	1.00	3.63	3.63
Geographical correlation	Combustion		Emission	air	Trisodium phosphate	kg/J	1.00	1.00	1.00	1.02	1.02
Geographical correlation	Combustion		Emission	air	PM10	kg/J	1.00	1.30	1.30	1.53	1.53
Geographical correlation	Combustion		Emission	air	Suspended solids, unspecified	kg/J	1.00	1.00	1.00	1.23	1.23
Geographical correlation	Combustion		Emission	air	PM10	kg/m	1.00	1.00	1.00	3.87	3.87
Geographical correlation	Combustion		Emission	air	Manganese	kg/kg	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Combustion		Emission	air	Dinitrogen	kg/J	1.00	9.62	9.62	8.54	8.54
Geographical correlation	Combustion		Emission	air	Methane	kg/kg	1.00	1.00	1.00	2.17	2.17
Geographical correlation	Combustion		Emission	air	Cadmium	kg/J	1.00	1.00	1.00	1.00	1.00
Geographical	Combustion		Emission	air	Dinitrogen	kg/kg	1.00	1.00	1.00	6.91	6.91

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
correlation											
Geographical correlation	Combustion		Emission	air	Carbon dioxide	kg/J	1.00	1.00	1.00	11.91	11.91
Geographical correlation	Combustion		Emission		Carbon dioxide	kg/kg	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Agriculture		Product exchange		Electricity	J/m2	1.00	1.00	1.00	1.78	1.78
Geographical correlation	Agriculture		Product exchange		Phosphorous pentoxide	kg/kg	1.00	1.00	1.93	1.23	1.05
Geographical correlation	Agriculture		Product exchange		Potassium oxide	kg/kg	1.00	1.00	1.09	2.94	2.75
Geographical correlation	Agriculture		Product exchange		Nitrogen fertilizer	kg/m2	1.00	1.00	1.00	4.45	4.45
Geographical correlation	Agriculture		Product exchange		Nitrogen fertilizer	kg/kg	1.00	1.00	6.21	2.28	2.40
Geographical correlation	Agriculture		Product exchange		Diesel	kg/kg	1.00	1.00	1.02	1.06	1.37
Geographical correlation	Agriculture		Product exchange		Electricity	J/kg	1.00	1.00	1.09	11.61	13.88
Geographical correlation	Agriculture		Product exchange		Seed	kg/m2	1.00	1.00	1.00	6.67	6.67
Geographical correlation	Manufacturing		Emission		unspecified PM	kg/kg	1.00	4.79	4.79	4.88	4.88
Geographical correlation	Manufacturing		Product exchange		Electricity	J/kg	1.00	1.01	1.01	1.06	1.06
Geographical correlation	Manufacturing		Emission		VOC	kg/kg	1.00	14.63	14.63	10.06	10.06
Geographical correlation	Manufacturing		Product exchange		Steam	J/kg	1.00	1.00	1.00	1.07	1.07
Geographical correlation	Manufacturing		Emission		unspecified PM	kg/m3	1.00	1.00	1.00	2.89	2.89
Geographical correlation	Manufacturing		Emission		Carbon monoxide	kg/kg	1.00	2.33	2.33	96.10	96.10
Geographical correlation	Manufacturing		Product exchange		Waste	kg/kg	1.00	1.21	1.21	17.54	17.54
Geographical correlation	Manufacturing		Emission		BOD	kg/kg	1.00	1.00	1.00	88.75	88.75
Geographical correlation	Manufacturing		Emission		Methanol	kg/kg	1.00	28.41	28.41	28.41	28.41
Geographical correlation	Manufacturing		Product exchange		Electricity	J/m3	1.00	1.00	1.00	21.66	21.66

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Geographical correlation	Manufacturing		Emission		Carbon dioxide, fossil	kg/kg	1.00	1.00	1.00	3.74	3.74
Geographical correlation	Manufacturing		Emission		Suspended solids, unspecified	kg/kg	1.00	1.00	1.00	8.43	8.43
Geographical correlation	Manufacturing		Emission		PM10	kg/kg	1.00	1.00	1.00	2.36	2.36
Geographical correlation	Manufacturing		Product exchange		Fuel	J/kg	1.00	1.00	1.00	2.05	2.05
Geographical correlation	Manufacturing		Emission		Formaldehyde	kg/kg	1.00	1.00	1.00	17.06	17.06
Geographical correlation	Manufacturing		Emission		Acids	kg/m3	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Manufacturing		Product exchange		Natural Gas	m3/m3	1.00	1.00	1.00	11.30	11.30
Geographical correlation	Manufacturing		Product exchange		Sodium component	kg/kg	1.00	1.00	1.00	1.81	1.81
Geographical correlation	Manufacturing		Emission		Nitrogen	kg/kg	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Manufacturing		Emission		Phenol	kg/kg	1.00	4.72	4.72	4.72	4.72
Geographical correlation	Manufacturing		Emission		Oil	kg/m3	1.00	1.00	1.00	1.00	1.00
Geographical correlation	Manufacturing		Emission		Formaldehyde	kg/kg	1.00	18.27	18.27	39.88	39.88
Geographical correlation	Manufacturing		Emission		Solids	kg/kg	1.00	1.00	1.00	123.03	123.03
Temporal correlation	Combustion		Emission	air	Nitrogen oxide	kg/kg	1.00	2.04	2.99	5.08	6.77
Temporal correlation	Combustion		Emission	air	Dinitrogen	kg/kg	1.00	3.89	3.89	3.89	6.15
Temporal correlation	Combustion		Emission	air	Carbon dioxide	kg/J	1.00	1.00	11.99	8.59	5.36
Temporal correlation	Combustion		Emission	air	NM VOC	kg/kg	1.00	1.13	1.98	1.98	1.98
Temporal correlation	Combustion		Emission	air	PM10	kg/J	1.00	1.27	1.09	1.36	1.42
Temporal correlation	Combustion		Emission	air	PM2.5	kg/J	1.00	1.36	1.17	1.27	1.27
Temporal correlation	Combustion		Emission	air	Dinitrogen	kg/J	1.00	1.00	1.08	1.43	1.43
Temporal correlation	Combustion		Emission	air	Nitrogen oxide	kg/J	1.00	1.17	1.03	1.24	1.18
Temporal correlation	Combustion		Emission	air	Carbon monoxide	kg/kg	1.00	1.74	1.05	1.26	1.75
Temporal correlation	Combustion		Emission	air	Sulfur oxide	kg/J	1.00	1.17	1.13	4.10	7.73

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Temporal correlation	Combustion		Emission	air	Methane	kg/J	1.00	1.00	1.12	1.13	1.05
Temporal correlation	Combustion		Emission	air	Carbon monoxide	kg/J	1.00	1.01	1.74	1.13	1.06
Temporal correlation	Agriculture		Product exchange		Electricity	J/m2	1.00	1.00	1.00	2.86	2.86
Temporal correlation	Agriculture		Emission	air	Methane	kg/head/year	1.00	4.19	4.19	6.40	6.40
Temporal correlation	Utilities		Emission	air	PM2.5	kg/J	1.00	1.00	6.18	6.18	6.18
Temporal correlation	Utilities		Emission	air	Carbon dioxide	kg/J	1.00	1.03	1.50	1.46	1.38
Temporal correlation	Utilities		Emission	air	PM10	kg/J	1.00	1.00	5.17	5.17	5.17
Temporal correlation	Utilities		Emission	air	Carbon monoxide	kg/J	1.00	1.00	1.28	13.64	13.64
Temporal correlation	Utilities		Emission	air	Formaldehyde	kg/J	1.00	1.00	1.00	18.00	18.00
Temporal correlation	Utilities		Emission	air	NM VOC	kg/J	1.00	1.00	1.07	1.13	1.13
Temporal correlation	Utilities		Emission	air	unspecified PM	kg/J	1.00	1.00	1.39	1.35	1.35
Temporal correlation	Utilities		Emission	air	Nitrogen oxide	kg/J	1.00	1.00	73.50	48.51	48.51
Temporal correlation	Utilities		Emission	air	Sulfur oxide	kg/J	1.00	1.00	2.19	4.79	4.79
Temporal correlation	Utilities		Emission	air	Dinitrogen	kg/J	1.00	1.00	1.33	1.67	1.67
Temporal correlation	Utilities		Emission	air	Methane	kg/J	1.00	1.00	1.27	1.54	1.54
Temporal correlation	Manufacturing		Emission	air	Carbon dioxide	kg/kg	1.00	1.03	110.88	57.90	55.88
Temporal correlation	Manufacturing		Emission	air	Dust	kg/kg	1.00	1.13	7.72	14.15	12.71
Temporal correlation	Manufacturing		Product exchange		Natural Gas	J/kg	1.00	16.03	11.52	7.28	7.28
Temporal correlation	Manufacturing		Emission	air	Nitrogen oxide	kg/kg	1.00	1.03	1.46	1.13	1.14
Temporal correlation	Manufacturing		Emission	air	Dinitrogen	kg/kg	1.00	1.07	1.59	1.59	1.31
Temporal correlation	Manufacturing		Emission	air	Fluorine	kg/kg	1.00	1.00	1.00	1.00	3.04
Temporal correlation	Manufacturing		Product exchange		Electricity	J/kg	1.00	73.89	7.44	7.45	7.37
Temporal correlation	Manufacturing		Emission	air	Sulfur oxide	kg/kg	1.00	1.51	1.15	1.35	1.88

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Temporal correlation	Manufacturing		Product exchange		Energy	J/kg	1.00	1.00	2.06	2.23	1.21
Temporal correlation	Manufacturing		Emission	air	Copper	kg/kg	1.00	1.00	2.70	2.70	2.70
Temporal correlation	Manufacturing		Emission	air	Mercury	kg/kg	1.00	1.00	5.21	2.90	2.90
Temporal correlation	Manufacturing		Emission	air	Chromium	kg/kg	1.00	1.00	1.69	1.69	1.69
Temporal correlation	Manufacturing		Emission	air	PM10	kg/kg	1.00	1.24	1.52	3.93	2.69
Temporal correlation	Manufacturing		Emission	air	Carbon monoxide	kg/kg	1.00	1.00	4.43	4.34	4.34
Temporal correlation	Manufacturing		Product exchange		Waste	kg/kg	1.00	1.00	9.17	7.32	7.32
Temporal correlation	Manufacturing		Emission	air	Methane	kg/J	1.00	1.80	1.80	1.80	1.80
Temporal correlation	Metal manufacturing		Product exchange		Energy	J/kg	1.00	1.00	2.55	2.55	2.55
Temporal correlation	Metal manufacturing		Product exchange		Silicon	kg/piece	1.00	1.84	2.36	2.36	2.36
Temporal correlation	Metal manufacturing		Product exchange		Lead	kg/piece	1.00	1.23	1.23	1.23	1.23
Temporal correlation	Metal manufacturing		Product exchange		Plastic	kg/piece	1.00	1.00	1.00	1.00	1.00
Temporal correlation	Metal manufacturing		Product exchange		Copper	kg/piece	1.00	1.00	1.00	1.00	1.82
Temporal correlation	Metal manufacturing		Product exchange		Steel	kg/piece	1.00	1.06	1.06	1.06	1.06
Temporal correlation	Metal manufacturing		Product exchange		Aluminium	kg/piece	1.00	1.30	1.30	1.30	1.30
Temporal correlation	Metal manufacturing		Product exchange			kg/piece	1.00	1.00	1.00	1.00	1.24
Temporal correlation	Transportation		Emission		Hydrocarbons	kg/m	1.00	1.24	1.45	1.43	1.31

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Temporal correlation	Transportation		Emission		Carbon monoxide	kg/m	1.00	1.23	1.35	1.28	1.11
Temporal correlation	Transportation		Emission		Nitrogen oxide	kg/m	1.00	1.14	1.13	1.09	1.02
Reliability	Combustion		Emission	air	Carbon dioxide	kg/J	1.00	1.20	1.79	1.79	18.77
Reliability	Combustion		Emission	air	Carbon dioxide	kg/m	1.00	1.00	1.00	1.00	1.00
Reliability	Combustion		Emission	air	Dioxin	kg/m3	1.00	1.00	1.00	1.00	1.91
Reliability	Combustion		Emission	air	Nitrogen oxide	kg/m3	1.00	1.00	1.00	1.00	1.14
Reliability	Combustion		Product exchange		Diesel	m3/m	1.00	1.10	1.10	1.10	1.10
Reliability	Agriculture		Emission	air	Methane	kg/head/year	1.00	1.00	1.35	1.35	6.40
Reliability	Agriculture		Product exchange		Seed	kg/m2	1.00	1.00	1.00	1.00	1.21
Reliability	Agriculture		Product exchange		Phosphorous pentoxide	kg/m2	1.00	1.00	1.00	1.00	1.06
Reliability	Agriculture		Product exchange		Herbicides	kg/m2	1.00	1.00	1.00	1.00	1.21
Reliability	Agriculture		Product exchange		Potassium oxide	kg/m2	1.00	1.00	1.00	1.00	1.55
Reliability	Agriculture		Product exchange		Nitrogen fertilizer	kg/m2	1.00	1.00	1.00	1.00	1.05
Reliability	Agriculture		Product exchange		Electricity	J/m2	1.00	1.00	1.00	1.00	1.78
Reliability	Manufacturing		Emission	water	Phosphorus	kg/kg	1.00	1.00	1.00	1.00	3.08
Reliability	Manufacturing		Emission	air	Carbon dioxide	kg/kg	1.00	1.00	2.47	2.47	1.22
Reliability	Manufacturing		Emission	air	NMVOC	kg/kg	1.00	1.00	54.59	54.59	38.44
Reliability	Manufacturing		Emission	water	Nitrogen fertilizer	kg/kg	1.00	1.00	1.00	1.00	2.00
Reliability	Manufacturing		Emission	air	PM10	kg/kg	1.00	1.00	27.74	27.74	14.50
Reliability	Manufacturing		Emission	air	VOC	kg/kg	1.00	1.00	1.69	1.69	1.69
Reliability	Manufacturing		Emission	air	Hydrogen fluoride	kg/kg	1.00	1.00	1.00	1.00	17.06
Reliability	Manufacturing		Emission	water	Suspended	kg/kg	1.00	1.00	1.25	1.25	1.25

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
	g				solids, unspecified						
Reliability	Manufacturing		Emission	water	COD	kg/kg	1.00	1.00	1.00	1.00	1.03
Reliability	Manufacturing		Emission	air	Dinitrogen	kg/kg	1.00	1.00	1.04	1.04	1.04
Reliability	Manufacturing		Emission	water	Phosphate	kg/kg	1.00	1.00	1.00	1.00	14.11
Reliability	Manufacturing		Emission	air	Copper	kg/kg	1.00	1.00	1.00	1.00	1.52
Reliability	Manufacturing		Emission	air	Chromium	kg/kg	1.00	1.00	1.00	1.00	1.28
Reliability	Manufacturing		Emission	air	Arsenic	kg/kg	1.00	1.00	1.00	1.00	7.80
Reliability	Manufacturing		Emission	air	Mercury	kg/kg	1.00	1.00	1.00	1.00	1.86
Reliability	Manufacturing		Emission	water	BOD	kg/kg	1.00	1.00	1.27	1.27	2.91
Reliability	Manufacturing		Emission	air	Nitrogen oxide	kg/kg	1.00	1.00	1.22	1.22	1.23
Reliability	Manufacturing		Emission	air	Dust	kg/kg	1.00	1.00	1.00	1.00	2.03
Reliability	Manufacturing		Emission	air	Carbon monoxide	kg/kg	1.00	1.00	1.12	1.12	1.05
Reliability	Manufacturing		Product exchange		Natural Gas	m3/kg	1.00	1.00	2.35	2.35	10.91
Reliability	Manufacturing		Product exchange		Heat	J/kg	1.00	1.00	1.00	1.00	33.38
Reliability	Manufacturing		Product exchange		Starch	kg/kg	1.00	1.00	1.00	1.00	5.23
Reliability	Manufacturing		Product exchange		Natural Gas	J/kg	1.00	1.00	1.00	1.00	2.24
Reliability	Manufacturing		Product exchange		Water	m3/kg	1.00	1.00	1.00	1.00	1.68
Reliability	Manufacturing		Product exchange		Energy	J/kg	1.00	1.00	1.00	1.00	1.95
Reliability	Manufacturing		Product exchange		Fuel	J/kg	1.00	1.00	1.00	1.00	9.77
Reliability	Manufacturing		Product exchange		Steam	J/kg	1.00	1.00	1.00	1.00	1.00
Reliability	Manufacturing		Product exchange		Pulp	kg/kg	1.00	1.00	1.00	1.00	1.69

Pedigree indicator	Industrial sector	Reference product	Type of flow	Compartment	Flow	Unit	UF1	UF2	UF3	UF4	UF5
Reliability	Manufacturing		Product exchange		Waste	kg/kg	1.00	1.00	1.20	1.20	9.17
Reliability	Metal manufacturing		Emission	water	Nickel	kg/kg	1.00	1.00	1.00	1.00	1.00
Reliability	Metal manufacturing		Emission	water	Copper	kg/kg	1.00	1.00	1.00	1.00	1.00
Reliability	Metal manufacturing		Product exchange		Energy	J/kg	1.00	1.00	1.00	1.00	1.88
Reliability	Metal manufacturing		Product exchange		Plastic	kg/piece	1.00	1.00	1.00	2.52	2.52

ONGLET *BUF* Données utilisées pour définir les fonctions de vraisemblance des facteurs d'incertitude de base

NAICS Level 4	Year	Country	Activity	Flow	Type of flow	Release compartment	Unit	Number of data	GSD2
1.A.3.b	2009	United Kingdom	Onroad activity	Carbon dioxide	Emission	Air	kg/m	1245	1.54
1.A.3.b	2009	United Kingdom	Onroad activity	Natural Gas	Product exchange		m3/m	22	1.69
1.A.3.b	2009	United Kingdom	Onroad activity	Diesel	Product exchange		m3/m	2328	1.70
1.A.3.b	2009	United Kingdom	Onroad activity	Petrol	Product exchange		m3/m	140	2.10
111110	1998	United States	Crop production	Seed	Product exchange		kg/m2	14	1.33
111110	1998	United States	Crop production	Gasoline	Product exchange		m3/m2	14	1.44
111110	1998	United States	Crop production	Oil	Product exchange		m3/m2	14	1.59
111110	1998	United States	Crop production	Phosphate fertilizer	Product exchange		kg/m2	14	1.95
111110	1998	United States	Crop production	Diesel	Product exchange		m3/m2	14	1.89
111110	1998	United States	Crop production	Potash	Product exchange		kg/m2	14	6.76

NAICS Level 4	Year	Country	Activity	Flow	Type of flow	Release compartment	Unit	Number of data	GSD2
111110	1998	United States	Crop production	Nitrogen fertilizer	Product exchange		kg/m2	14	3.63
111110	1998	United States	Crop production	Electricity	Product exchange		J/m2	10	57.12
111110	1998	United States	Crop production	LPG	Product exchange		m3/m2	11	34.32
112130	2010	Switzerland	Manure management	Methane	Emission	Air	kg/head/year	6	7.18
112990	2010	Switzerland	Manure management	Methane	Emission	Air	kg/head/year	2	10.59
112120	2010	Switzerland	Manure management	Methane	Emission	Air	kg/head/year	2	64.17
115111	1995	United States	Cotton ginning	PM2.5	Emission	Air	kg/kg	3	5.13
115111	1995	United States	Cotton ginning	PM2.5-10	Emission	Air	kg/kg	3	14.09
115111	1995	United States	Cotton ginning	PM10	Emission	Air	kg/kg	3	18.01
115114	1995	United States	Postharvest crop activities	PM10	Emission	Air	kg/kg	6	18.79
115114	1995	United States	Postharvest crop activities	PM2.5-10	Emission	Air	kg/kg	6	145.65
115114	1995	United States	Postharvest crop activities	PM2.5	Emission	Air	kg/kg	6	61.34
212325	2010	Spain	Clay mining	Waste	Product exchange	Air	kg/kg	2	1.62
212391	1995	United States	Potash mining	PM2.5	Emission	Air	kg/kg	2	51.67
212391	1995	United States	Potash mining	PM2.5-10	Emission	Air	kg/kg	2	73.64
212391	1995	United States	Potash mining	PM10	Emission	Air	kg/kg	2	95.71
213113	1995	United States	Coal mining - support activities	PM10	Emission	Air	kg/kg	4	174.84
213113	1995	United States	Coal mining - support activities	PM2.5	Emission	Air	kg/kg	4	122.29
213113	1995	United States	Coal mining - support activities	PM2.5-10	Emission	Air	kg/kg	4	172.41
213115	1995	United States	Mining - support activities	PM2.5	Emission	Air	kg/kg	16	118.00
213115	1995	United States	Mining - support activities	PM10	Emission	Air	kg/kg	16	132.13
213115	1995	United States	Mining - support activities	PM2.5-10	Emission	Air	kg/kg	16	127.36
221115	2001	Denmark	Wind power production	Nitrogen oxide	Emission	Air	kg/J	2	254.34
322121	1995	Netherlands	Pulp and paper	BOD	Emission	water	kg/kg	2	2.13

NAICS Level 4	Year	Country	Activity	Flow	Type of flow	Release compartment	Unit	Number of data	GSD2
			production						
322121	1995	Netherlands	Pulp and paper production	COD	Emission	water	kg/kg	2	2.13
322121	1995	Netherlands	Pulp and paper production	Nitrogen fertilizer	Emission	water	kg/kg	2	7.00
325110	2000	Netherlands	Organic chemicals production	VOC	Emission	Air	kg/kg	3	3.83
325110	2000	Netherlands	Organic chemicals production	Nitrogen oxide	Emission	Air	kg/kg	2	66.15
325180	2003	United States	Inorganic chemicals	Fuel	Product exchange		J/kg	14	13.43
325180	2003	United States	Inorganic chemicals	Steam	Product exchange		J/kg	32	26.11
325180	2003	United States	Inorganic chemicals	Electricity	Product exchange		J/kg	43	375.13
325199	1995	United States	Basic organic chemical manufacturing	PM10	Emission	Air	kg/kg	4	41.27
325199	1995	United States	Basic organic chemical manufacturing	PM2.5-10	Emission	Air	kg/kg	4	55.79
325199	1995	United States	Basic organic chemical manufacturing	PM2.5	Emission	Air	kg/kg	4	72.72
325199	2003	United States	Organic chemicals	Fuel	Product exchange		J/kg	5	6.00
325199	2003	United States	Organic chemicals	Electricity	Product exchange		J/kg	43	42.07
325199	2003	United States	Organic chemicals	Steam	Product exchange		J/kg	42	27.22
327110	2010	Spain	Clay - ceramics manufacturing	Copper	Emission	Air	kg/kg	2	1.11
327110	2010	Spain	Clay - ceramics manufacturing	Chromium	Emission	Air	kg/kg	2	2.31
327110	2010	Spain	Clay - ceramics manufacturing	Diesel	Product exchange		m3/kg	3	5.90
327110	2010	Spain	Clay - ceramics manufacturing	Natural Gas	Product exchange		J/kg	3	11.42
327110	2010	Spain	Clay - ceramics manufacturing	Nitrogen oxide	Emission	Air	kg/kg	3	4.21
327110	2010	Spain	Clay - ceramics manufacturing	Sulfur oxide	Emission	Air	kg/kg	3	13.34
327110	2010	Spain	Clay - ceramics manufacturing	Water	Product exchange		m3/kg	3	10.17

NAICS Level 4	Year	Country	Activity	Flow	Type of flow	Release compartment	Unit	Number of data	GSD2
327110	2010	Spain	Clay - ceramics manufacturing	Mercury	Emission	Air	kg/kg	2	7.91
327110	2010	Spain	Clay - ceramics manufacturing	Arsenic	Emission	Air	kg/kg	2	10.34
327110	2010	Spain	Clay - ceramics manufacturing	Lead	Emission	Air	kg/kg	2	12.55
327110	2010	Spain	Clay - ceramics manufacturing	PM10	Emission	Air	kg/kg	2	18.91
327110	2010	Spain	Clay - ceramics manufacturing	Electricity	Product exchange		J/kg	2	19.50
327110	2010	Spain	Clay - ceramics manufacturing	Carbon monoxide	Emission	Air	kg/kg	3	57.41
327110	2010	Spain	Clay - ceramics manufacturing	Hydrogen fluoride	Emission	Air	kg/kg	3	32.42
327110	2010	Spain	Clay - ceramics manufacturing	Waste	Product exchange		kg/kg	6	241.59
327310	2004	Estonia	Cement production	Phenol	Emission	water	kg/kg	2	3.58
327410	2006	Spain	Lime production	Energy	Product exchange		J/kg	5	1.52
331313	1995	United States	Aluminium manufacturing	PM2.5-10	Emission	Air	kg/kg	2	9.00
331313	1995	United States	Aluminium manufacturing	PM10	Emission	Air	kg/kg	2	9.33
331313	1995	United States	Aluminium manufacturing	PM2.5	Emission	Air	kg/kg	2	9.60
331511	2001	Italy	Iron Foundry	Benzo-fluoranthene	Emission	Air	kg/m3	2	13.00
331513	1995	United States	Steel foundries	PM2.5	Emission	Air	kg/kg	3	44.30
331513	1995	United States	Steel foundries	PM2.5-10	Emission	Air	kg/kg	3	40.71
331513	1995	United States	Steel foundries	PM10	Emission	Air	kg/kg	3	38.71
335911	1995	United States	Battery manufacturing	PM10	Emission	Air	kg/piece	5	83.70
335911	1995	United States	Battery manufacturing	PM2.5-10	Emission	Air	kg/piece	5	99.50
335911	1995	United States	Battery manufacturing	PM2.5	Emission	Air	kg/piece	5	148.49
562213	1995	United States	Waste incinerator	PM10	Emission	Air	kg/kg	2	1.77
562213	1995	United States	Waste incinerator	PM2.5-10	Emission	Air	kg/kg	2	1.96
562213	1995	United States	Waste incinerator	PM2.5	Emission	Air	kg/kg	2	2.08

ONGLET *Lognormal_Assumptions* Tester l'hypothèse de log-normalité pour la distribution des facteurs d'incertitude

The assumption of lognormality is tested on the uncertainty data used to derive likelihood functions for both the additional and the basic uncertainties. For the additional uncertainty, the assumption is tested for the generic factors of the four pedigree indicators: reliability, temporal correlation, geographical correlation and further technological correlation and for the scores 2 to 5 (the additional uncertainty factor for the completeness indicator is not derived from data, the assumption cannot be tested). For the basic uncertainty factors, the number of data obtained and useable to derive the likelihood function is often 1 or 2. Only the number of data useable to define the basic uncertainty likelihood functions for the demand of thermal energy, electricity, semi-finished products, materials, waste for the processes and the PM2.5 emissions to air for processes are obtained based on more than 10 uncertainty figures. The assumption of lognormality is so tested on these two uncertainty factors.

The lognormal distribution is tested using the Kolmogorov-Smirnov test.

<i>Assessed parameters</i>	<i>Number of data in the sample</i>	<i>D from Kolmogorov Smirnov test</i>	<i>Critical value for the Kolmogorov Smirnov test with a level alpha = 0.05</i>	<i>Test passed?</i>
BUF_Thermal	13	0.1497	0.3614	OK
BUF PM2.5	16	0.2628	0.3723	OK
Reliability UF2	46	0.5357	0.1963	NO
Reliability UF3	46	0.3656	0.1963	NO
Reliability UF4	46	0.3528	0.1963	NO
Reliability UF5	46	0.2261	0.1963	NO
Temporal correlation UF2	52	0.3482	0.1848	NO
Temporal correlation UF3	52	0.2221	0.1848	NO
Temporal correlation UF4	52	0.1695	0.1848	OK
Temporal correlation UF5	52	0.1599	0.1848	OK
Geographical correlation UF2	59	0.3905	0.1752	NO
Geographical correlation UF3	59	0.3718	0.1752	NO
Geographical correlation UF4	59	0.1635	0.1752	OK
Geographical correlation UF5	59	0.1641	0.1752	OK
Further technological correlation UF2	72	0.3804	0.1597	NO
Further technological correlation UF3	72	0.2812	0.1597	NO

<i>Assessed parameters</i>	<i>Number of data in the sample</i>	<i>D from Kolmogorv Smirnov test</i>	<i>Critical value for the Kolmogorov Smirnov test with a level alpha = 0.05</i>	<i>Test passed?</i>
Further technological correlation UF4	72	0.2231	0.1597	NO
Further technological correlation UF5	72	0.1968	0.1597	NO

CONCLUSION

A majority of parameters failed at the Kolmogorov-Smirnov test. Nonetheless, when plotting these parameters, they all follow an assymetric distribution. When assessing these parameters using the *fitdistrplus* package in R⁶, the representation using the Cullen and Frey graphic (that plots the skewness against the kurtosis) shows that the parameters might follow a Beta distribution. Given this observation and the assymetry of the data, the lognormal distribution is kept to model uncertainty factors.

Furthermore, what is important in this study is the mean of the distribution rather than the distribution itself.

ONGLET *Posterior distribution* Paramètres des distributions log-normales à posteriori

All the posterior distributions are supposed to be lognormal distributions.

The parameters used to define lognormal distributions are μ and σ , respectively the lognormal mean and the lognormal standard deviation.

The values in the cells are the couples (μ ; σ)

1st assumption: σ_{IUF} is defined based on the 95th confidence interval (see Figure 6-2)

⁶ Marie Laure Delignette-Muller and Christophe Dutang, *fitdistrplus*: An R package for fitting distributions, *Journal of statistics software* (2015) 64(4) 1-34

2nd assumption: $\sigma_{2\text{ UF}}$ is defined based on the 66th confidence interval (see **Erreur ! Source du renvoi introuvable.**)

	FIRST ASSUMPTION			
	<i>UF2</i>	<i>UF3</i>	<i>UF4</i>	<i>UF5</i>
RELIABILITY				
Generic	(0.01 ; 4.42e-3)	(0.18 ; 0.08)	(0.22 ; 0.07)	(0.85 ; 0.14)
Agriculture	(0.00 ; 0.00)	(0.05 ; 0.04)	(0.07 ; 0.04)	(0.45 ; 0.19)
Combustion	(0.05 ; 0.03)	(0.12 ; 0.08)	(0.17 ; 0.07)	(0.49 ; 0.27)
Manufacturing	(0.00 ; 0.00)	(0.18 ; 0.09)	(0.22 ; 0.08)	(0.94 ; 0.18)
Metal Manufacturing	(0.00 ; 0.00)	(0.00 ; 0.00)	(0.19 ; 0.08)	(0.40 ; 0.19)
COMPLETENESS				
Generic	(0.023 ; 5.57e-6)	(0.050;3.66e-6)	(0.098 ; 1.22e-5)	(0.49 ; 4.09e-3)
TEMPORAL CORRELATION				
Generic	(0.08 ; 0.04)	(0.44 ; 0.10)	(0.79 ; 0.12)	(0.90 ; 0.12)
Agriculture	(0.03 ; 0.05)	(0.12 ; 0.13)	(0.56 ; 0.22)	(1.40 ; 0.07)
Combustion	(0.07 ; 0.05)	(0.23 ; 0.12)	(0.53 ; 0.16)	(0.52 ; 0.29)
Utilities	(0.00 ; 0.00)	(0.19 ; 0.13)	(0.60 ; 0.22)	(0.89 ; 0.25)
Manufacturing	(0.04 ; 0.05)	(0.31 ; 0.12)	(0.74 ; 0.18)	(0.88 ; 0.23)
Metal Manufacturing	(0.06 ; 0.04)	(0.19 ; 0.10)	(0.27 ; 0.12)	(0.39 ; 0.08)
Transportation	(0.15 ; 0.02)	(0.23 ; 0.06)	(0.23 ; 0.07)	(0.14 ; 0.04)
GEOGRAPHICAL CORRELATION				
Generic	(0.13 ; 0.06)	(0.20 ; 0.08)	(0.85 ; 0.14)	(0.82 ; 0.14)
Agriculture	(0.00 ; 0.00)	(0.09 ; 0.10)	(0.43 ; 0.18)	(0.43 ; 0.17)
Combustion	(0.08 ; 0.06)	(0.10 ; 0.07)	(0.50 ; 0.16)	(0.49 ; 0.15)
Manufacturing	(0.08 ; 0.08)	(0.14 ; 0.10)	(0.66 ; 0.19)	(0.64 ; 0.18)
FURTHER TECHNOLOGICAL CORRELATION				
Generic	(0.17 ; 0.06)	(0.41 ; 0.10)	(0.66 ; 0.12)	(0.80 ; 0.07)
Agriculture	(0.00 ; 0.00)	(0.27 ; 0.11)	(0.45 ; 0.15)	(0.70 ; 0.08)
Combustion	(0.04 ; 0.01)	(0.04 ; 0.03)	(0.40 ; 0.12)	(0.63 ; 0.08)
Utilities	(0.00 ; 0.00)	(0.25 ; 0.11)	(0.47 ; 0.14)	(0.69 ; 0.07)
Chemical manufacturing	(0.14 ; 0.08)	(0.36 ; 0.11)	(0.70 ; 0.14)	(0.79 ; 0.08)

SECOND ASSUMPTION				
	<i>UF2</i>	<i>UF3</i>	<i>UF4</i>	<i>UF5</i>
RELIABILITY				
Generic	(0.01 ; 4.42e-3)	(0.23 ; 0.10)	(0.26 ; 0.10)	(0.93 ; 0.15)
Agriculture	(0.00 ; 0.00)	(0.04 ; 0.04)	(0.05 ; 0.04)	(0.47 ; 0.22)
Combustion	(0.05 ; 0.04)	(0.13 ; 0.10)	(0.15 ; 0.09)	(0.60 ; 0.41)
Manufacturing	(0.00 ; 0.00)	(0.27 ; 0.14)	(0.27 ; 0.12)	(1.11 ; 0.20)
Metal Manufacturing	(0.00 ; 0.00)	(0.00 ; 0.00)	(0.20 ; 0.14)	(0.39 ; 0.22)
COMPLETENESS				
Generic	(0.023 ; 5.58e-6)	(0.053 ; 1.16e-5)	(0.11 ; 4.35e-5)	(0.80;8.34e-3)
TEMPORAL CORRELATION				
Generic	(0.16 ; 0.07)	(0.64 ; 0.12)	(0.90 ; 0.13)	(0.96 ; 0.13)
Agriculture	(0.04 ; 0.10)	(0.17 ; 0.26)	(0.98 ; 0.32)	(1.18 ; 0.35)
Combustion	(0.15 ; 0.08)	(0.39 ; 0.17)	(0.66 ; 0.19)	(0.78 ; 0.22)
Utilities	(0.00 ; 0.00)	(0.37 ; 0.22)	(1.03 ; 0.31)	(1.21 ; 0.33)
Manufacturing	(0.08 ; 0.09)	(0.65 ; 0.20)	(1.06 ; 0.23)	(1.18 ; 0.23)
Metal Manufacturing	(0.10 ; 0.06)	(0.25 ; 0.12)	(0.28 ; 0.13)	(0.39 ; 0.13)
Transportation	(0.17 ; 0.03)	(0.25 ; 0.07)	(0.23 ; 0.08)	(0.13 ; 0.07)
GEOGRAPHICAL CORRELATION				
Generic	(0.22 ; 0.09)	(0.29 ; 0.09)	(1.16 ; 0.16)	(1.15 ; 0.16)
Agriculture	(0.00 ; 0.00)	(0.18 ; 0.16)	(0.77 ; 0.25)	(0.77 ; 0.25)
Combustion	(0.12 ; 0.08)	(0.13 ; 0.08)	(0.75 ; 0.20)	(0.75 ; 0.20)
Manufacturing	(0.21 ; 0.13)	(0.31 ; 0.16)	(1.29 ; 0.27)	(1.29 ; 0.27)
FURTHER TECHNOLOGICAL CORRELATION				
Generic	(0.29 ; 0.09)	(0.62 ; 0.13)	(0.82 ; 0.15)	(0.95 ; 0.11)
Agriculture	(0.00 ; 0.00)	(0.44 ; 0.19)	(0.54 ; 0.25)	(0.73 ; 0.15)
Combustion	(0.04 ; 0.01)	(0.03 ; 0.03)	(0.39 ; 0.16)	(0.54 ; 0.12)
Utilities	(0.00 ; 0.00)	(0.35 ; 0.17)	(0.55 ; 0.22)	(0.69 ; 0.11)
Chemical manufacturing	(0.35 ; 0.14)	(0.69 ; 0.19)	(1.08 ; 0.22)	(1.01 ; 0.15)

ONGLET *Data_Gypsum* Données utilisées pour définir les fonctions de vraisemblance utilisées à la section 6.2.3.3

Name	Unit	Subgroup	UF_1	UF_2	UF_3	UF_4
Joint tape	kg/m2	1	1.00	1.05	1.24	1.24
Block of wood	kg/m2	2	1.00	3.43	3.63	3.63
Block of gypsum	kg/m2	3	1.00	1.00	2.44	2.44
Coating	kg/m2	4	1.00	1.91	1.98	2.47
PE film	kg/m2	5	1.00	2.18	2.01	2.74
Gypsum board	kg/m2	6	1.00	2.17	2.70	2.72
Screws	kg/m2	7	1.00	1.92	1.57	1.56